**VILNIUS UNIVERSITY**
**LIFE SCIENCES CENTER**

IEVA JASKOVIKAITĖ

Molecular biotechnology study program

**Master thesis**

# THE ROLE OF DNA TOPOLOGY IN CRISPR-CAS9 NUCLEASE SPECIFICITY

Student: Ieva Jaskovikaitė

Supervisor: Dr Stephen Knox Jones Jr.

Vilnius, 2023

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

4mC – 4-methylcytosine

5caC – 5-carboxylcytosine

5fC – 5-formulcytosine

5hmC – 5-hydroxymethylcytosine

5mC – 5-methylcytosine

6mA – 6-methyladenine

CRISPR – clustered regularly interspaced short palindromic repeats

CRISPR-Cas – CRISPR associated protein

CRISPRi – CRISPR interference

crRNA – CRISPR RNA

dCas9 – nuclease-dead Cas9

DNMT – DNA methyltransferase

Fis – factor for inversion stimulation

FluoRBT – fluorescent rotor bead tracking

ghmC – glucosyl-hydroxymethylcytosine

gRNA – guide RNA

H3K27m3 – trimethylation of histone 3 lysine 27

H3K9m2 – dimethylation of histone 3 lysine 9

hmC – hydroxymethylcytosine

H-NS – Histone-like nucleoid-structuring protein

HU – histone-line protein from *E. coli* strain U93

IHF – integration host factor

KRAB – Krüppel-associated box domain

Lrp – leucine-responsive regulatory protein

PAM – protospacer adjacent motif

RBT – rotor bead tracking

REP – repetitive extragenic palindromic

RNase III – ribonuclease III

SAM – S-adenosyl-L-methionine

SARS-CoV-2 - severe acute respiratory syndrome coronavirus 2

sgRNA – single guide RNA

SMC – Structural maintenance of chromosomes protein

tracrRNA – *trans*-activating CRISPR RNA

# INTRODUCTION

Bacteriophages infect bacteria and destroy them in their quest to replicate (Kasman and Porter, 2023). To defend against bacteriophages, bacteria have evolved several defense systems, one of which is an adaptive immune system – clustered regularly interspaced short palindromic repeats, or CRISPR in short (Jansen et al., 2002). For this system to work, CRISPR associated (CRISPR-Cas) proteins recognize a short protospacer adjacent motif (PAM) sequence in bacteriophage DNA, and then cut out the sequence next to it - a protospacer (Heler et al., 2015). By adding this protospacer to its genome, bacterium retains a memory of the infection and prevents the next one by making a guide RNA (gRNA) using it as template (Barrangou et al., 2007; Brouns et al., 2008).

Recently, CRISPR-Cas proteins, most popularly CRISPR-Cas9 from *Streptococcus pyogenes*, have been repurposed for gene editing in eukaryotic cells (Adli, 2018; Doudna and Charpentier, 2014). There, CRISPR-Cas9 can be programed to find a desired DNA target (also referred to as the on-target) with the help of a target-complementary single guide RNA (sgRNA), where the nuclease can cleave and allow DNA to be edited (Jinek et al., 2012). However, the nuclease also edits untargeted places, called off-targets (Hsu et al., 2013). To better understand and prevent this unanticipated activity, one should understand how the eukaryotic environment affects this prokaryotic nuclease. One of the challenges associated with this environmental change is the differences between prokaryotic and eukaryotic DNA topology, which the nuclease is now exposed to.

One topological feature that separates prokaryotes from eukaryotes is DNA methylation. Prokaryotes methylate adenine and cytosine (Seong et al., 2021), while eukaryotes mainly methylate DNA on cytosine (Parashar et al., 2018; Schmitz et al., 2019). Methylation on prokaryotic DNA is crucial for determining self and non-self, as this pattern is recognized by restriction-modification systems that help to prevent bacteriophage infections (Bikard and Marraffini, 2012). Conversely, eukaryotes methylate their genomes as means to control their gene expression, a process called epigenetic gene control (de Mendoza et al., 2020). In 2018, the Kallimasioti-Pazi and colleagues observed that when CRISPR-Cas9 system is programmed towards an epigenetically imprinted DNA, the nuclease cleaves the DNA more slowly (Kallimasioti-Pazi et al., 2018). However, the magnitude of this effect is unclear for CRISPR-Cas9 on- versus off-target cleavage kinetics.

Another important difference investigated in this project is supercoiling. In prokaryotes, DNA is circular (Bendich and Drlica, 2000) and negatively supercoiled for compaction (Joyeux, 2015). Positive

supercoils appear during transcription (Chedin and Benham, 2020), but apart from that are not tolerated and are immediately fixed (Schvartzman et al., 2019). In contrast, eukaryotic DNA can be positively and negatively supercoiled, depending on the transcription process. However, in a eukaryotic environment, both supercoils briefly co-exist (Corless and Gilbert, 2017). In 2020, Ivanov and colleagues showed that CRISPR-Cas9 cleaves DNA that partially matches its sgRNA programming (i.e., off-targets) more readily when supercoiled than when relaxed (Ivanov et al., 2020). However, the magnitude of this effect on target specificity and cleavage kinetics for Cas9 is still unknown.

The studies mentioned in this section begin to illustrate the ways DNA topology can impact CRISPR-Cas9 activity. This study seeks more insight on how DNA topology affects CRISPR-Cas9 nuclease specificity. First, this study aims to investigate the changes in CRISPR-Cas9 cleavage kinetics for on- and off-targets when a linear DNA target is either unmodified or carries adenine or cytosine methylation. The impact of methylation at every location within the DNA target, as well as the only adenine within the PAM, was tested on cleavage by CRISPR-Cas9. The second part of this project investigates how CRISPR-Cas9 cleavage rate changes in response to differential DNA supercoiling for both on- and off-targets. A target library containing matched and mispaired sequences (compared to the sgRNA) was inserted into a plasmid and then either supercoiled or relaxed by different topoisomerases before exposure to CRISPR-Cas9 cleavage. The data collected in this study will provide more insights into the specificity of the enzyme by providing cleavage rate data for methylated and negatively supercoiled on- and off-targets. For the supercoiling assay, a biophysical model will be created that will allow a better prediction of cleavage results for an *in vivo* application of a desired target.

# AIM AND OBJECTIVES

Project aim:

Evaluate if the topology (methylation and supercoiling) of on-target and off-target DNAs impacts their recognition and cleavage by CRISPR-Cas9 from *S. pyogenes*.

Project objectives:

1. Assess if adenine or cytosine methylation of linear DNA targets changes Cas9 on-target cleavage rates compared to unmethylated linear target DNAs.
2. Assess if adenine or cytosine methylation of linear DNA targets changes Cas9 off-target cleavage rates compared to unmethylated linear target DNAs.
3. Determine if negatively supercoiling DNA promotes Cas9 off-target cleavage when compared to relaxed DNA.

# 1. LITERATURE REVIEW

## 1.1 Introduction to CRISPR-Cas system

### The new bacterial immune system

The history of clustered regularly interspaced short palindromic repeats (CRISPR) (Jansen et al., 2002) is colorful enough that a poetic scientist could classify it into historical periods spanning antiquity and middle ages to the contemporary period (Barrangou and Horvath, 2017). Regardless of one's affinity for extravagant naming, the first traces of CRISPR touched by man were at the end of the 20$^{th}$ century. Back then, a strange genetic anomaly was first described: palindromic sequences of 35 nucleotides long could be found repeated many times over in *Escherichia coli* and *Salmonella typhimurium* genomes, so-called repetitive extragenic palindromic (REP) sequences (Stern et al., 1984). After a few years, a group was investigating the isozyme conversion of alkaline phosphatase in *E. coli*. After sequencing *iap* – the gene potentially encoding the proteolytic enzyme responsible for such a conversion – they saw the same unusual sequence repeats, which could form secondary structures (Ishino et al., 1987). Yet another group showed that REPs stabilized their neighboring mRNA sequence, and the resulting construct regulated bacterial protein expression (Newbury et al., 1987). As years passed, researchers found more sequences with dyad symmetry in *Shigella dysenteriae* (Nakata et al., 1989) and even in archaea such as *Haloferax mediterranei* (Mojica et al., 1993) and *Haloferax volcanii* (Mojica et al., 1995). In 2002, bioinformaticians finally identified four CRISPR associated (cas) genes. The *cas1-4* genes were found adjacent to the CRISPR locus, but only two of the four had a possible function: Cas3, a superfamily 2 helicase homologue, could modify the DNA, while Cas4, which had similarities to RecB exonuclease, might bind intact and cleaved DNA (Jansen et al., 2002). Arguably the biggest break-through was in 2005, when three research groups figured out that the spacers – the varying sequences between the palindromic repeats – originated from bacteriophages (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). These spacers are responsible for the bacteria's resistance to infection by bacteriophage with matching sequence, while new spacers are acquired upon new infection by unfamiliar bacteriophages (Barrangou et al., 2007). Soon after, Brouns *et al* (2008) showed that the palindromes, together with the spacers, make specific guide RNAs (gRNA) that guide Cas proteins to sites in a prophage's genome (Brouns et al., 2008), where cleavage inactivates it (Garneau et al., 2010).

As the years went by, more CRISPR systems were found that differed from one another in their overall composition. These systems were divided into two classes that differed in the way that their

effector complexes are organized – class I have several subunits making up the effector complex, while class II only have one effector protein (Ishino et al., 2018). Additional subdivisions defined as many as 6 types and 33 subtypes, which further captured the variability among CRISPR-Cas systems (Janik et al., 2020). Types can differ based on the Cas proteins that are present in the CRISPR locus – the presence of Cas3 is a signature of type I; the infamous Cas9 is present in type II systems; Cas10 is a member of type III, while the absence of Cas1 and Cas2 defines type IV CRISPR-Cas systems (Ishino et al., 2018). Different subtypes capture the "small" differences: *e.g.*, type II systems, are subtyped into II-A, II-B, and II-C. They have identical Cas9, Cas1, and Cas2 proteins, but II-A also has a Csn2, II-B replaces Csn2 with Cas4, while II-C omits a fourth Cas protein (Shmakov et al., 2017).

Having many different CRISPR-Cas systems is not only beneficial for bacteria but this versatility can also be employed by scientists outside of it. CRISPR-Cas9 helps to perform genetic knockout screens that identify what cancer and stem cells need to survive and genes that confer resistance to drug treatment (Shalem et al., 2014). CRISPR-Cas12a and Cas13, which both specifically cleave their intended target and then any available substrate (Kordyś et al., 2022), underlie disease diagnostics for SARS-CoV-2 (Broughton et al., 2020), Zika and Dengue viruses (Gootenberg et al., 2017). But for gene editing, CRISPR-Cas9 remains the most widely used nuclease due to its single multidomain effector, a short (5'-NGG-3', where N is any nucleotide) protospacer adjacent motif (PAM) requirement (Adli, 2018), and ease of target reprogramming via changing the gRNA sequence (Doudna and Charpentier, 2014). For these reasons, this review concentrates on the type II CRISPR-Cas9 system and its mechanisms of action.

A closer look at how CRISPR-Cas9 works

When it comes to the main working mechanisms of CRISPR-Cas9, there are three steps to neutralize the foreign genetic material (summarized in Figure 1.1) (Janik et al., 2020). The very first step is adaptation. Type II systems, to which CRISPR-Cas9 belongs (Shmakov et al., 2017), rely on several Cas proteins for the execution of this phase – Cas1, Cas2, and Cas9 – in addition to the *trans*-activating CRISPR RNA (tracrRNA). Upon infection, the Cas9-tracrRNA complex recognizes foreign DNA (a protospacer) next to Cas9's PAM, which then can be excised (Heler et al., 2015). Then, the Cas1-Cas2 complex recognizes a five-nucleotide sequence separating the palindromic repeats of the CRISPR locus from its upstream leader sequence. The latter sequence is responsible for orienting new spacers so that they can be integrated by the order of infection, which increases the bacteria's fitness. Thus, Cas1-Cas2 integrate the acquired protospacer right next to the leader, expanding the CRISPR array (Figure 1.1A)

(McGinn and Marraffini, 2016). An additional protein, Csn2, is also required for spacer acquisition *in vivo* and makes a complex with the aforementioned proteins, but its exact role remains to be elucidated (Wilkinson et al., 2019).

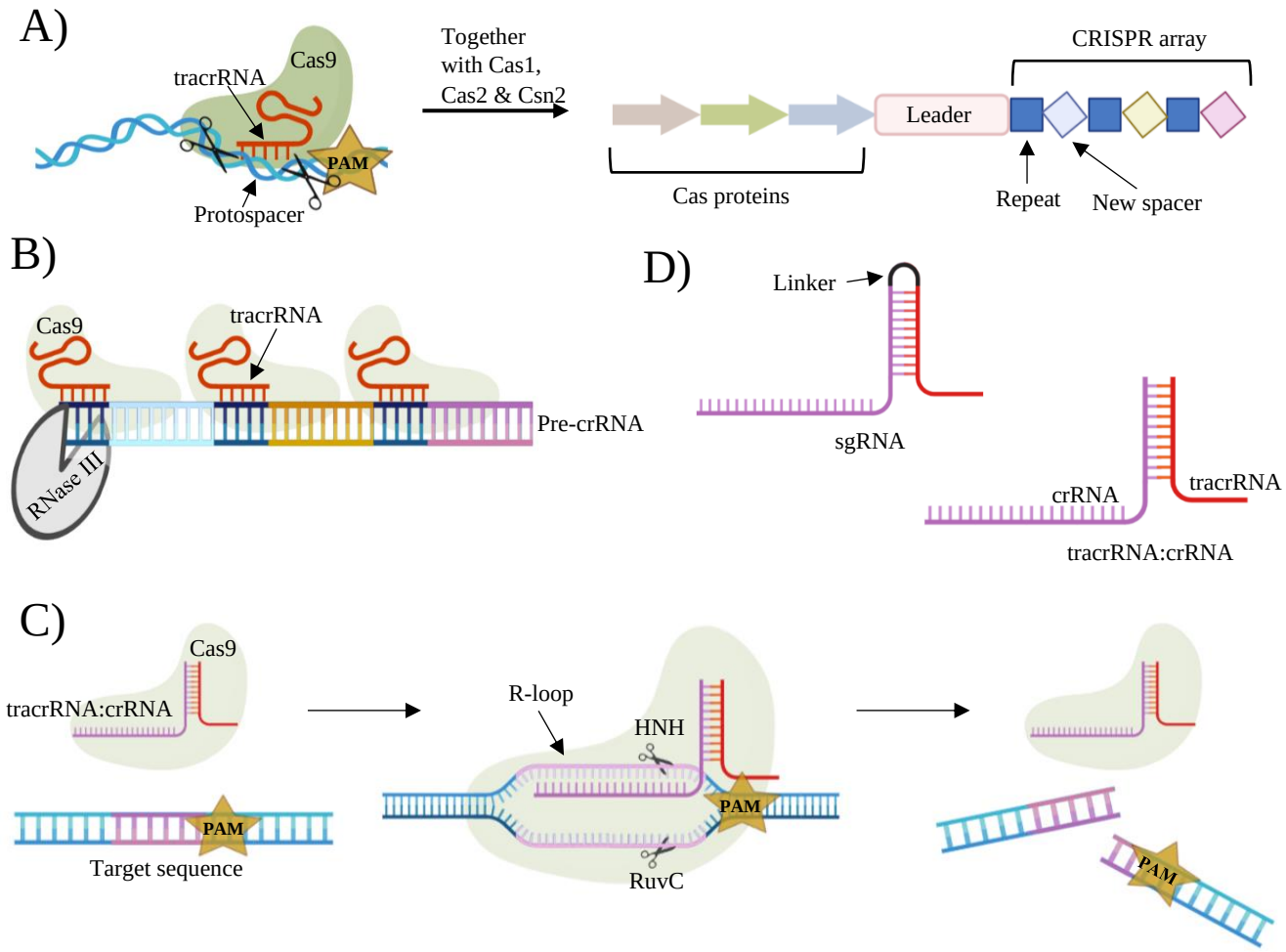


**Figure 1.1 The three stages of CRISPR-Cas9 bacterial immunity.** Panel A) illustrates the process of adaptation, where Cas9 helps to recognize the target next to the PAM sequence, while additional Cas proteins help to insert the new spacer in the CRISPR array. Panel B) schematically shows the tracrRNA:crRNA biogenesis process with the help of RNase III. In panel C) during interference, CRISPR-Cas9 recognizes PAM-proximal target sequence, forms R-loop and executes cleavage with RuvC and HNH nuclease domains before dissociating from the inactivated invader sequence. D) illustrates the difference of engineered sgRNA. Made with BioRender.com

After a spacer sequence is integrated into the CRISPR array, the second step is RNA guide biogenesis (Figure 1.1B). This step involves several different components, namely tracrRNA, Cas9, host ribonuclease III (RNase III), and the spacer with palindromic repeats next to it – pre-crRNA. First, the unprocessed tracrRNA, which can be as long as 171 base pairs, gets processed to ~75 base pairs before

it can bind pre-crRNA repeated sequences via base complementarity. This complex is then bound by Cas9 and processed by RNase III. Following another processing event, the pre-crRNA is processed to mature crRNA of 39-42 nucleotide in length, and only then this tracrRNA:crRNA construct becomes a functional guide for Cas9 (Deltcheva et al., 2011). The 5' end of tracrRNA and 3' end of crRNA are held together by 22 paired nucleotides, but it is possible to engineer a single guide RNA (sgRNA) chimera, by joining crRNA with a tracrRNA via a linker (Figure 1.1D) (Jinek et al., 2012).

Once a known genetic aggressor enters the bacteria, the final step – interference – takes place. In type II systems, Cas9 binds and cleaves this invading DNA sequence with the help of tracrRNA:crRNA, which has formed a complex with Cas9 during crRNA biogenesis. The Cas9-RNA ribonucleoprotein is directed to the correct target with the help of a PAM sequence, which is recognized as a double-stranded DNA (dsDNA) though Cas9's PAM interacting domain (Nishimasu et al., 2014). The PAM's sequence is read from the non-target strand, where deviations in the sequence demolish this type II systems' ability to execute target cleavage, thus supporting the importance of the PAM location on the non-target strand (Jinek et al., 2012). The DNA melting starts at the PAM, after it is recognized. Melting continues toward the end of the target sequence, forming a gRNA-DNA hybrid with the non-target strand getting flipped out to the other side, a structure called R-loop (Figure 1.1 C). This formation of the R-loop depends on base pairing between gRNA and target strand and is especially mismatch-sensitive in the seed region – up to 12 PAM-proximal base pairs (Sternberg et al., 2014). If the Mg ions are present in the environment (Jinek et al., 2012) and the substrate is DNA (as Cas9 is unable to cleave RNA substrates (Gasiunas et al., 2012)), then the HNH nuclease domain cleaves the target strand, while RuvC nuclease domain takes care of the non-target strand (Figure 1.1C). Cas9 cleaves its target strand three base pairs upstream of the PAM sequence (Jiang and Doudna, 2017), while the non-target strand can be cleaved three to four bases upstream the PAM (Zeng et al., 2018). This type of cleavage has a speed of about 0.3 to 1 $min^{-1}$ and leaves blunt ends on the target DNA. The acceptable DNA topology for Cas9 is versatile as it can cleave both linear and supercoiled targets, which then allows Cas9 to cleave the same foreign DNA several times should it have several matching targets (Jinek et al., 2012).

Is CRISPR-Cas9 specific?

The interference step has a lot of requirements for the target specificity – it has to be proximal to the correct PAM sequence (Nishimasu et al., 2014) and it has to complement the 20 nucleotide gRNA (Sternberg et al., 2014). This makes it look like it does not leave a lot of space for any mismatches to be

tolerated, however this would not be beneficial for bacteria. Bacteriophages cannot replicate on their own and rely completely on infected bacteria's machinery to synthesize new proteins for the assembly of the progeny (Weigel and Seitz, 2006). Bacteria have several defense systems that are able to recognize the invading genetic material and neutralize it (Yuan et al., 2023). This ability to recognize invading bacteriophages puts them on constant pressure to slightly change their genetic code sequence in hopes this abolishes recognition by the bacteria (Naureen et al., 2020). Because of this, if CRISPR-Cas9 could only cleave the sequences that perfectly matched the PAM and sgRNA-complementarity requirements, it would not be a good defense system for long.

In addition to the well-established PAM sequence of NGG (where N can be any nucleotide), it has been shown that another sequence – NAG – is also recognized by Cas9. This sequence is approved much slower than the canonical NGG PAM – NAG-proximal targets are cleaved five times slower (Hsu et al., 2013). Single mismatches within the target sequence are also tolerated, but the extent is dependent on the position within the sequence itself, as the further from the seed region, the less of an impact a mismatch has. Additionally, the amount of enzyme is important for the mismatch tolerance, as using Cas9 in excess reduces its sensitivity to mismatched PAM-proximal sequences (Pattanayak et al., 2013). More than one mismatch between the DNA and sgRNA is possible, as even six nucleotide mismatches also allow cleavage *in vivo*, indicating that Cas9 interrogation of DNA is not as sequence-constrained as it would be desired (Tsai et al., 2015). Even deletions and insertions in the DNA sequence (relative to the gRNA), forming a gRNA or DNA bulge, respectively, do not abolish cleavage (Lin et al., 2014).

CRISPR-Cas9's off-target activity poses a great threat of introducing unwanted and possibly dangerous edits and thus limits CRISPR-Cas9's *in vivo* application range. One way around this issue is to use an engineered Cas9 variant that is made specifically to evade the better part of off-targets by, for example, having a reduced cleavage rate (Liu et al., 2020). However, in order to know the best way to evade off-targets, the mechanisms allowing them must be fully understood with the help of future work.

### 1.2 Introduction to DNA topology – methylation and supercoiling

#### Bacteriophages

As the CRISPR-Cas system defends against bacteria-invading viruses, one should consider what bacteriophage genomes look like. The genetic material that gets injected into the bacteria is rather versatile – it can be single or double-stranded DNA or RNA (Hatfull and Hendrix, 2011), which is why

different nucleases can cleave either DNA, such as Cas9 (Gasiunas et al., 2012), or RNA, such as Cas13a (Liu et al., 2017). Due to the exceptionally small size of bacteriophages, their DNA usually does not have any modification on them (for example, methylation), which serves as an identification advantage for the bacteria (Marks and Sharp, 2000). The overall compaction of double-stranded genetic material in bacteriophages acquires the form of a toroid (Hud, 1995), or coaxial coils (Jiang et al., 2006), and even knots (Marenduzzo et al., 2009), which all help to compactly store DNA and easily unwind upon infection.

Procaryotes

Bacteria can have their DNA sequences with additional "on-top" modifications – methylation patterns. There are three different families of enzymes capable of methylating the DNA in bacteria – Dam, Dcm, and CcrM methyltransferases (Casadesús, 2016). Two nucleobases can be methylated in bacteria – adenine and cytosine – with the help of the methyl group donor S-adenosyl-L-methionine (SAM) (Seong et al., 2021). Methylation on the 6N position of adenine (6mA) is the predominant form of methylation in bacteria (Figure 1.2), while methylation of positions 4N and 5C on cytosine (4mC and 5mC, respectively) is less abundant (Beaulaurier et al., 2019). The methylation patterns concerning the DNA strands can also vary, with either one strand carrying the methyl group (hemimethylation) or both (Adhikari and Curtis, 2016). The ability to methylate their genomes comes as a great advantageous option for the bacteria – it enables innate immunity against bacteriophage infection by a restriction-modification system. These systems consist of a methylase and a restriction endonuclease that work hand in hand to create a methylation pattern on the bacterial DNA. The pattern is recognized by the cleaving enzyme as "self and uncleavable". Bacteriophages lack this modification and thus are exposed to degradation by the restriction-modification system (Bikard and Marraffini, 2012). However, this does not hold true for all methyltransferases, as some do not associate with any nucleases (the so-called "orphan"

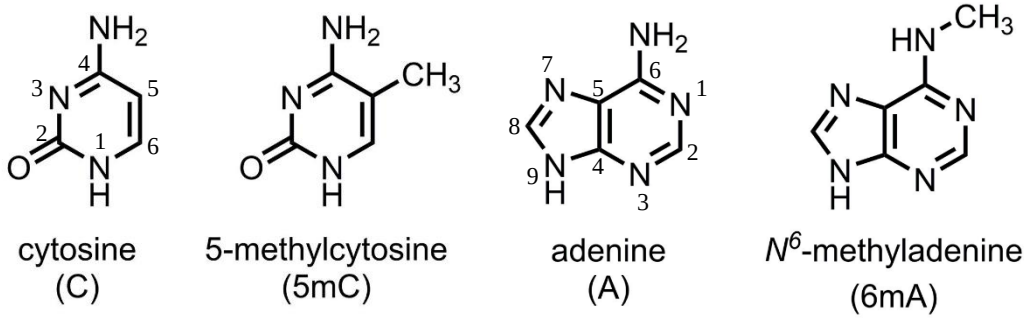


**Figure 1.2 Examples of adenine and cytosine methylation.** Modified from Shi et al., 2017

methyltransferases) and do not provide innate immunity, but rather have regulatory roles. For example, CcrM is expressed only during chromosome replication and thus, it is important in the regulation of the cell cycle (Casadesús, 2016). Dcm, another lone methyltransferase, has been shown to generate sites susceptible to mutations, contributing to the plasticity of the genetic material as well as the possible evolution of bacteria (Beaulaurier et al., 2019).

With regards to the DNA topology in bacteria, it is rather simple – there is a circular DNA molecule that usually compacts into one chromosome (as sometimes there is even more than one chromosome!), and can be accompanied by one to several plasmids that carry additional beneficial genes (Bendich and Drlica, 2000). As opposed to the eukaryotic DNA compartmentalization into a membranous nucleus, bacteria have their genetic material unseparated from the cytoplasm – it is located in the nucleoid, which takes up to 25% of the whole cell volume.

The bacterial DNA, just as the eukaryotic DNA, is a B-form DNA, and makes a turn every 10.5 base pairs (Joyeux, 2015), while gyrases further negatively supercoil the DNA (Kuzminov, 2014) and keep it at the superhelical density of -0.05 (Joyeux, 2015). Meanwhile, reverse gyrases positively supercoil thermophilic prokaryotes' DNA (Kuzminov, 2014). However, these positive or negative supercoiling directionalities become easily subject to the DNA polymerase during transcription. As the polymerase moves along the DNA molecule, it applies strong torsional stress on the DNA, resulting in the formation of positive supercoils in front of the replication fork, leaving the usual negative supercoils behind it (Chedin and Benham, 2020). But bacterial gyrases work hard to maintain the crucial negative supercoiling of the DNA, thus eliminating positive supercoils introduced during transcription (Schvartzman et al., 2019). However, supercoiling alone is not enough to sufficiently condense the DNA: nucleoid-associated proteins help. The histone-like nucleoid structuring protein (H-NS), structural maintenance of chromosomes complexes (SMC), and leucine-responsive regulatory protein (Lrp) help to bridge the surrounding DNA duplexes. Other proteins, such as integration host factor (IHF), histone-like protein from *E. coli* strain U93 (HU), and factor for inversion stimulation (Fis) can perform the bending of prokaryotic DNA (Luijsterburg et al., 2006).

Eukaryotes

Just as the bacteria, eukaryotic cells also have post-replication modifications on their DNA, but they differ from the aforementioned ones (summarized in Table 1.1 at the end of this section). The methyl group is placed from the cofactor SAM on the $5^{th}$ carbon of the cytosine ring. This is executed by DNA

methyltransferases (DNMTs), which can either place the methyl groups in a *de novo* fashion or maintain the already existing pattern (Schmitz et al., 2019). DNMT3A and DNMT3B are responsible for placing methyl groups at novel positions, thus creating *de novo* methylation patterns within the genome. Another methyltransferase – DNMT1 – works to maintain the methylation pattern on both strands equally after replication (Greenberg and Bourc'his, 2019). The nucleotides that are susceptible to such chemical modifications are located within the long stretches of CG dinucleotide repeats, known as CpG islands. Evolutionary, not every single CpG island is methylated; the ones exempt are in promoter regions, which serve to allow the transcription machinery to bind and begin transcription (Dhar et al., 2021). The methylated cytosine can have some variation in higher eukaryotes – it can be either 5-methylcytosine (5mC) (Figure 1.2), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Shi et al., 2017). These forms do not show any differences when it comes to base pairing but might affect the strength of the formed hydrogen bonds and their interactions with DNA-binding proteins (Ren et al., 2018). A methyl group on adenine (6mA) is also possible, but is more common in mRNA and is only present at minuscule amounts in DNA (Parashar et al., 2018). It is rather well-known that methylated DNA sequences in eukaryotic DNA are involved in gene silencing and X-chromosome inactivation. However, adding methyl groups on DNA does not necessarily result in transcription silencing, since methyl groups can physically repel the binding of transcription factors. It can have an adverse effect and lead to the increased binding affinity of several transcription factors, thus leading to increased transcription (de Mendoza et al., 2020). This especially holds true for introns, as the methyl groups there have a favorable effect on the initiation of transcription by the histone modifiers and chromatin remodelers (Dhar et al., 2021). Lastly, DNA methylation is not the only chemical modification that can control gene expression patterns in eukaryotes. Histone proteins can also harbor methyl and acetyl groups, further contributing to the repression or expression of genes (Kouzarides, 2002).

The compaction of eukaryotic genome has some similarities with prokaryotic genome (see Table 1.1 for a summary). Eukaryotic DNA, located in the double-membrane nucleus, is linear. Telomeres signify the ends of each linear molecule (Bendich and Drlica, 2000). When the genetic material is condensed, it makes up a set number of chromosomes – a karyotype – which depends on the organism, usually ranging from 10 to 100 (Kuzminov, 2014). Histone proteins (H2A, H2B, H3, H4) together make a nucleosome, a structure around which wraps the DNA. The DNA wrapped around many histones is separated by roughly 10-60 base pairs (the so-called linker DNA) and is referred to as beads-on-a-string or the 10nm fiber. The binding of additional non-histone proteins, such as cohesins, establishes the final

chromatin structure, which is negatively charged (Maeshima et al., 2020). The eukaryotic DNA can be over- and under-wound, and the cause of such topology is usually a DNA-binding protein, such as DNA/RNA polymerase. These enzymes, just as in prokaryotic DNA, leave negative supercoils behind the replication fork, while making positive supercoils right ahead. Nucleosomes, as another DNA binding protein example, wrap 147 base pairs of DNA in a negative supercoiling manner (Corless and Gilbert, 2017), while the linker DNA is left unconstrained. Should the supercoiling be unfavorable for the cell, it can be released with the help of topoisomerases, which exert their function on the linker DNA. There are two topoisomerases that execute this – topoisomerase I (nicks a single DNA strand and rotates it; is mostly present where transcription takes place) and topoisomerase II (performs a double-stranded cut). Usually, the under-wound state of DNA is more favorable for transcription and R-loop formation, compared to over-wound DNA, which prevents polymerase movement across the DNA due to high torsional stress (Corless and Gilbert, 2016). The distribution of under- and over-wound DNA in eukaryotes might depend on the need to express the genes encoded within the DNA, thus permitting the existence of both at the same time (Corless and Gilbert, 2017).

**Table 1.1 Summary contrasting the prokaryotic and eukaryotic DNA topology** (continues on the next page)

| | **Prokaryotes** | **Eukaryotes** |
|---|---|---|
| | **Methylation** | |
| **Methylated nucleobases** | Adenine (most prevalent; 6mA) & cytosine (4mC, 5mC) | Cytosine (5mC; 5hmC; 5fC; 5caC) & sometimes adenine (6mA) |
| **Methylation result** | Self and non-self differentiation; chromosome replication | Epigenetic gene expression control |
| **Methylating enzymes** | Dam, Dcm, CcrM | DNMT1, 3A, 3B |
| **Additional modifications?** | No | Methylation and acetylation of histones |
| | **Compaction of the genome** | |
| **DNA location** | Nucleoid in the cytoplasm (relies on phase separation) | In the nucleus (two membranes separate from cytoplasm) |

| | | |
|---|---|---|
| **Supercoiling** | Mainly negative, positive only ahead of the replication machinery | Locus-specific positive and negative (required for gene expression); positive ahead of the replication machinery |
| **DNA associated proteins** | Help to further compact the genome; H-NS, SMC, Lrp, IHF, HU, Fis | Help to further compact the genome; nucleosomes and e.g., cohesins |
| **DNA supercoiling enzymes** | Gyrase | Topoisomerase I and II |

### 1.3 DNA methylation and CRISPR-Cas9 activity

#### Does DNA methylation have an effect?

Methylation is an example of chemical DNA markers that help bacterial nucleases avoid DNA that is meant to be kept intact (Bikard and Marraffini, 2012). Therefore, it makes sense that the CRISPR-Cas9 system should also be tested for susceptibility to such modifications, as it is also a nuclease of bacterial origin. Quite a few studies were performed that looked at the interactions between Cas9 and methylated DNA. For example, Hsu *et al* in 2013 performed an *in vitro* cleavage assay with methylated and naked pUC19 plasmids to test if there were any cleavage differences between these two substrates. The targets were designed in a way that the methyl groups could cover either the sgRNA matching sequence or the PAM region. After performing the experiment, they observed cleavage of both plasmids, which lead them to conclude that at least *in vitro,* the position of the methyl group within the target or PAM sequence was insufficient to prevent cleavage (Hsu et al., 2013). However, as only *in vitro* experiments do not show the whole picture, the authors asked if this held true with the methylated targets in the *in vivo* setting. To test this, they selected a human gene target *SERPINB5,* which harbors many methylation sites, and they used 3 sgRNAs that targeted different regions within the locus. The outcome of this experiment was in accordance with the one from the *in vitro* study – the locus was successfully edited despite having many epigenetic marks (Hsu et al., 2013). This research was rather a conclusive one, as it involved both the *in vitro* and *in vivo* settings, but there were many different variables (e.g., the sgRNA number which was too low to test all different positions of the methyl groups within the given target sequence) that could have tipped the scales to the negative answer.

As the years followed, other research teams also tried their luck with DNA methylation and Cas9 cleavage experiments. However, this time a bigger emphasis was placed on the *in vivo* setting, as the results could be more directly applied to genome editing with CRISPR-Cas9. Fujita and colleagues in 2016 investigated CRISPR-Cas9 ability to edit a gene located in alleles with different methylation statuses. They chose *p16INK4a* from the human colorectal carcinoma cell line HCT116 for this purpose. One copy of *p16INK4a* is heavily methylated on the DNA as well as histones (H3K9m2) and silenced due to chromatin compaction, while the other allele bears the unmethylated and expressed *p16INK4a* (Fujita et al., 2016). They tested a total of 3 sgRNAs that targeted an increasing number of methylated CpG sequences (up to 5) in their targets. The methyl groups within such CpG sequences were also present in the PAM and the PAM-proximal region. The results showed that CpG methylation did not decrease the editing efficiencies to a significantly lower level for two sgRNA targets, but did prevent one sgRNA (called sgRNA_rig3) from editing its target. However, such a difference between sgRNA_rig3 and the other two sgRNAs was highly unlikely to have happened due to methylation – sgRNA_rig3 only had 3 methylated CpG sites, while one of the other targets harbored more (Fujita et al., 2016). When they assayed the binding capabilities of these sgRNAs coupled with catalytically inactive "dead" Cas9 (dCas9), the results showed that the sgRNA_rig3 preferred the naked DNA rather than the methylated one. Because such preference of sgRNA_rig3 for both binding and cleavage was not observed in other sgRNAs, it could be possible that this preference relates to chromatin accessibility rather than methylation. This is also supported by the *in vitro* dCas9 binding to the sgRNA_rig3 target, which showed no difference between the naked and methylated DNA once the compaction into higher chromatin orders was not present (Fujita et al., 2016). Therefore, the study suggests that the differences observed in cleavage of methylated versus naked DNA most likely are not a result of methyl groups impeding the Cas9 activity, but may depend on additional factors, such as nucleosome positioning in compacted chromatin.

A study by Kallimasioti-Pazi and colleagues in 2018 also looked at the effect the epigenetic imprinting (methylation) could have on the cleavage by Cas9. The novelty of the paper included additional investigation of the possible imprinted allele cleavage outcomes, such as insertions or deletions when compared to the expressed allele. To test this, they investigated several CpG sites in mouse embryonic stem cells that were known to be methylated in the maternal allele – *KvDMR1, Impact,* and *Inpp5f_v2*. Then, these loci were exposed to either high, intermediate or low concentrations of Cas9. The mutagenicity frequency and the subsequent outcome in maternal (epigenetically imprinted) and

paternal (expressed) alleles were monitored for up to 96 hours. The cleavage rate, measured as the percentage of the target cleaved over time, only had a significant difference at intermediate and high concentrations of Cas9, while low concentrations seemed to achieve comparable results. In the first 24 hours, intermediate and high concentrations of the enzyme in the maternal allele achieved 13% to 23% of targets cleaved, respectively, while in the paternal allele these percentages went up to 33% and 78%, respectively. Interestingly, the mutagenicity rates were not different on either allele, which lead them to conclude that it is the cleavage rate that changes when the target of interest is imprinted (Kallimasioti-Pazi et al., 2018).

The result of methylation is what matters

Could these observed differences be attributed to the methyl groups that silenced the maternal allele? This could be the case, as it is known that some other nucleases from bacteria can indeed recognize methylated DNA and thus avoid cleaving it (Bikard and Marraffini, 2012). But does Cas9 actually respond to such DNA modification, or does it respond to more compact chromatin, completely disregarding the additional methyl groups, as was suggested by the results of Fujita and colleagues in 2016? One way to investigate if the methyl groups or their resulting heterochromatin formation is the real cause of observed *in vitro* differences in cleavage would be to have heterochromatin which is not compacted by DNA methylation, but by, for example, histone modifications. Polycomb Repressive Complex can place three methyl groups on the $27^{th}$ lysine on H3 (H3K27me3), which results in the compaction of chromatin into heterochromatin (Simon and Kingston, 2009). With the help of this complex Daer and colleagues in 2017 investigated how different chromatin states – open, partially closed, and closed – affected Cas9 cleavage in the HEK293 cell line. The results showed that the biggest slowing of Cas9-mediated editing was observed when they targeted the first 150 base pairs of the transcription start site in the compacted chromatin. Moving away from the transcription start site still had an inhibitory effect on the nuclease activity, but some edits prevailed and were still possible. Interestingly, one sgRNA showed no decrease in editing efficiency when was introduced to the heterochromatin target, maybe due to uneven distribution of the compacted chromatin state (Daer et al., 2017). Similar results by Chen and colleagues in 2016 indicated that it is the heterochromatin that impedes Cas9 function. They chose a different chromatin compaction method in human cells – they used Krüppel-associated box domain (KRAB) as a transcriptional repressor which induces chromatin compaction through H3K9me3 (Chen et al., 2016). These selected examples act as a suggestion that

maybe methylation is not the reason Cas9 has different cleavage rates *in vivo* (Kallimasioti-Pazi et al., 2018). It is more likely that the compact form of DNA serves as a physical barrier because once this component is eliminated, the nuclease does seem to function with the same efficiency as it does with naked DNA (Fujita et al., 2016). But does this mean that no chemical DNA modification affects Cas9 cleavage?

Outside-of-the-box thinking, DNA methylation, and clever bacteriophages

Bacteriophages and bacteria are in a constant arms race, and one way to avoid being recognized by the host is to add additional chemical groups to the DNA, such as various methyl (e.g. hydroxymethylcytosine – hmC) and glucosyl (glucosyl-hydroxymethylcytosine – ghmC) groups (Lehman and Pratt, 1960). Such ghmC can prevent restriction systems from recognizing foreign DNA, but it is unknown if the same applies to the CRISPR-Cas system. To test this, several versions of bacteriophages were used to challenge Cas9 – T4 bacteriophage with hmC, T4 with ghmC, and RB49 with no DNA modifications. In all three cases Cas9 was successfully able to recognize and cleave the invading phage DNA, thus showing that methylation (neither on cytosine nor adenine) and additional glycosylation is not a hindrance (Yaung et al., 2014).

Does this mean that CRISPR-Cas9 is insensitive to chemical DNA modifications and all bacteriophages' trouble to methylate their genomes was in vain? Bryson *et al* seemed to have a different perspective on the same question, which seems to be solely due to having gRNAs that are not connected by the linker (for the differences refer back to Figure 1.1D). The investigators also looked at T4 bacteriophage and its different mutants which allowed them to have either hmC or ghmC. They also used several gRNAs targeting different sequences which had varying levels of modified cytosines but excluded every sequence that also carried methylation on adenine, to have study-specific results only. After challenging the CRISPR-Cas9 system with these bacteriophages, the authors found that ghmC modification of either DNA strand was significantly effective at stopping Cas9 from recognizing and cleaving the invading T4 bacteriophage. Only one gRNA targeting the least modified DNA sequence showed some prevention against the bacteriophage infection. Similar results were observed for the hmC version of T4, as it was also able to prevent degradation. However, the same gRNA that worked for hmC, showed even better activity levels for ghmC. This suggests that additional chemical modifications work more efficiently, but do not necessarily completely halt Cas9 cleavage (Bryson et al., 2015).

Vlot *et al* in 2018 partially supported the findings from Bryson and colleagues, as they also found that ghmC inhibits Cas9 ability to prevent T4 infection. Vlot *et al* (2018) also noted that hmC bacteriophages that escaped Cas9 had single mutations in the seed region, which might have been the cause of their escape, not the methylation of the cytosine. They also investigated why ghmC prevents Cas9 cleavage, and it turns out that it impairs the nuclease affinity to bind due to the introduction of several steric clashes as compared to the one with an unmodified target. Additionally, it was shown that such inhibition is only effective when located in the target sequence, as ghmC in the PAM sequence did not prevent cleavage (Vlot et al., 2018). These examples illustrate that a single methyl group on cytosine is insufficient to protect from Cas9 degradation. However, as it was already there with the hopes of preventing the degradation by the restriction-modification system, it was a great place for further additions of chemical groups, which proved effective against the adaptive immune system that is the CRISPR-Cas9.

### 1.4 DNA supercoiling and CRISPR-Cas9 activity

#### Torque influence on R-loop formation

CRISPR-Cas9 requires a lot of physical contact with DNA before it can execute cleavage. Therefore, it is important to know the precise mechanisms that take place when the enzyme binds its substrate and starts R-loop base pairing. Cas9 can bend the DNA it is interrogating into a V shape upon the recognition of the PAM (Cofsky et al., 2022). Such physical distortion of the DNA helps unwind the PAM-proximal sequence to begin the search for matches. The enzyme does not check the whole sequence at the same time as PAM-proximal DNA testing is carried out by flipping one base pair at a time toward the gRNA. Because flipping bases introduces additional underwinding (Cofsky et al., 2022), it could suggest why Cas9 should have trouble binding and cleaving positive supercoils: under higher torsional stress, it should be even more difficult to melt the target by flipping bases and forming a fully hybridized R-loop. The opposite assumption goes for negative supercoiling when the DNA is underwound, which should further help to facilitate the opening of the R-loop.

Torque dependency was tested by Ivanov and colleagues in 2020, where they used a rotor bead tracking (RBT) approach to apply both negative and positive torque to target DNA. Either torque is achieved as one rotates a magnetic bead attached to the desired DNA. When the DNA was overwound, dCas9 was not able to melt the DNA duplex and form the R-loop, but as soon as the DNA became underwound, R-loop formation was easily observable. They also determined that the R-loop formed

within the first 100 ms of nuclease binding (Ivanov et al., 2020). The introduction of positive torques stimulated the release of the DNA by dCas9, which led to subsequent collapse of the R-loop. The R-loop forms through an intermediate stage, which is when the seed region melts. This happens during the unwinding and dissociating phases of dCas9. These intermediates were important for defining why the PAM-distal mismatches are more tolerated than the PAM-proximal ones. If there was a mismatch in the intermediate stage forming sequence, then this stage was quick to collapse under both positive and negative torques, but if the mismatch was further away from the seed region, then the Cas9 dwelled longer and more stably on the intermediate stage, allowing full R-loop formation at more negative torque (Ivanov et al., 2020).

The same torque-dependency applied for Cas9 cleavage – if the target was linear (aka relaxed), then the cleavage could not happen when there were mismatches between the target and gRNA. But if the target was negatively supercoiled, then both PAM-proximal and distal mismatches still allowed for some of the cleavages to be executed, as negative torque shifts the energy landscape of Cas9 binding to favor the transition to the open, ready-to-cleave stage (Ivanov et al., 2020). This has some negative consequences for the applications for the off-targets of Cas9 in eukaryotic genome editing, which is mainly negatively supercoiled (Corless and Gilbert, 2017). However, as the positive supercoils seemed to have an inhibitory effect on the Cas9 cleavage, it could be that a temporary creation of positive supercoils on some off-targets would protect them from being cut. The other side of the coin would be if the required edit would be present in the part that is located within the positively supercoiled part of the DNA. Then, an opposite effect would be needed to somehow uncoil the required site and potentially transfer the positive supercoils to any nearby off-targets. Nevertheless, it could be that even if the negative supercoiling allows for off-target cleavage, the rate at which such targets are degraded is slow enough so that choosing a nuclease with lower off-target tolerance could achieve the desired result.

### Nucleosomes obstruct Cas9 activity

Eukaryotic genomes have additional structural factors than prokaryotic genomes, and one of them are nucleosomes, which help to compact the DNA (Maeshima et al., 2020). In 2016, Horlbeck and colleagues investigated human cell lines *in vivo* using CRISPR interference (CRISPRi), a CRISPR system based gene silencing tool, where dCas9 binds to the gene of interest, physically preventing its transcription (Zhang et al., 2021). The results showed that CRISPRi was much more active at sites with low nucleosome occupancy, thus suggesting that histone octamers prevent Cas9 access. Additionally, the

sgRNAs which should have been highly active only showed the expected level of activity in nucleosome-unoccupied DNA targets (Horlbeck et al., 2016). The results *in vivo* strongly pointed to nucleosomes having an inhibitory effect on Cas9's activity. It was further investigated in an *in vitro* setting with target DNA wrapped around mouse histone. Both binding and cleavage experiments with this DNA showed a halting of Cas9 activity if the target was in contact with the nucleosome. The authors proposed that the main reason behind the observed results was due to the nucleosomes introducing significant steric hindrances for Cas9 (Horlbeck et al., 2016). There were other studies investigating the nucleosome effect for Cas9 cleavage *in vivo*. Yarrington and colleagues in 2018 looked at the possibility for Cas9 to cleave targets over time at the *HO* promoter from *Saccharomyces cerevisiae*. They found that nucleosome-free sites were cleaved more over time, while nucleosome-occupied targets were resistant to cleavage. However, *in vivo* nucleosomes "breathe" by making more loose contacts with the DNA, allowing some Cas9 cleavage, which was observed in the study (Yarrington et al., 2018). Treating nucleosome-bound DNA with Reb1, a protein that releases DNA from nucleosomes, was found to restore cleavage to that of naked DNA, illustrating the importance of nucleosome obstruction of cleavage. They also confirmed that dCas9 binding of nucleosome-bound DNA can be restored to naked-DNA levels when Reb1 is introduced (Yarrington et al., 2018).

Although the studies discussed so far have been conducted *in vivo* and *in vitro*, other papers have concentrated more extensively on just the *in vitro* observations. It is possible to investigate the effect that nucleosomes have on CRISPR-Cas9 function by focusing on targets within a short dsDNA sequence with high nucleosome affinity, called the Widom 601 positioning sequence (Lowary and Widom, 1998). One study by Isaac and colleagues in 2016, showed that if Cas9 targeted sequences closer or on the histone octamer, the cleavage rates were significantly reduced, while Cas9 targeting nucleosome flanking sequences showed rates that were close to those observed for naked DNA (Isaac et al., 2016). Additionally, the authors were curious about which step of the Cas9 mechanism nucleosomes interfere with. The authors measured the ability of dCas9 to bind 601 sequences with and without nucleosomes, which were designed to have no flanking DNA. The experiments indicated that while the nucleosome-unbound 601 sequence showed high dCas9 binding, the nucleosome-bound sequence had undetectable dCas9 binding. Nevertheless, their results indicated that the nuclease was still able to interrogate the DNA for PAM sequences, but it was unable to form a stable R-loop (Isaac et al., 2016). As the 601 sequence is characterized by very strong binding to the nucleosome (Lowary and Widom, 1998), it was also desirable to test if the observed results held true for the DNA that is allowed to "breathe" – slightly

dissociate from the nucleosome. Thus, a nucleosome with a 5S rRNA gene sequence from *Xenopus borealis* was used to perform the same tests. The results showed that nucleosome breathing allows Cas9 to achieve much higher cleavage rates at all sites except for the ones in the middle of the nucleosome (Isaac et al., 2016). Another study by Hinz and colleagues in 2015 showed that the position of PAM is also crucial for Cas9's ability to cleave targets associated with the nucleosomes. The study was carried out *in vitro* with a 601 nucleosome positioning sequence which had two sgRNA-complementary targets. They were designed so that the target sequences would be at the nucleosome boundary, but one would have its PAM on the linker side and the other would have its PAM on the nucleosome side. The target with its PAM on the nucleosome had a severe reduction of cleavage (more than 80% compared to the naked DNA), compared to the one with its PAM outside the nucleosome (about 19% reduction compared to naked DNA) (Hinz et al., 2015). This supports the observations that PAM interrogation still happens even on tightly bound nucleosomes (Isaac et al., 2016), but the R-loop formation is severely halted by nucleosome presence (Hinz et al., 2015).

## 1.5 Conclusion

CRISPR-Cas9 is a valuable gene editing tool sourced from bacterial efforts to prevent destruction by bacteriophages (Barrangou et al., 2007; Brouns et al., 2008; Garneau et al., 2010). However, once applied to a new environment – the eukaryotic genome –some things need to be taken into consideration for efficient gene editing: CRISPR-Cas9 encounters new DNA topological features, such as nucleosome wrapping (Maeshima et al., 2020). The reviewed literature seems to agree that these features can interfere with the nuclease's activity and impact cleavage both *in vivo* (Horlbeck et al., 2016; Yarrington et al., 2018) and *in vitro* (Hinz et al., 2015; Isaac et al., 2016). DNA methylation sometimes help bacteria differentiate between self and non-self DNA (Bikard and Marraffini, 2012), but do not provide the same protection from the CRISPR-Cas9 system (Fujita et al., 2016; Hsu et al., 2013; Kallimasioti-Pazi et al., 2018). However, DNA methylation in eukaryotic cells also helps to compact DNA (de Mendoza et al., 2020), and the resulting condensation of chromatin does have a negative effect on the cleavage rate by Cas9 (Chen et al., 2016; Daer et al., 2017). Thus, methylation still affects the activity of this nuclease but not through a direct route of introducing steric clashes. When investigating different torques within the supercoiled DNA, it is now known that R-loop formation is indeed torque-dependent as negative torque promotes promiscuous Cas9 activity (Ivanov et al., 2020), which should be taken into consideration when trying to target specific genes *in vivo*.

# 2. MATERIALS AND METHODS

## 2.1 Linear DNA design

A small set of five linear DNAs, named methylation targets 1-5 (targets M1-5), was created to test methylation effect on Cas9 cleavage kinetics. The targets were designed to have sliding sequences of two repeating motifs (TAA and AAT with unvarying CG motif between them; underlined in Table 2.1) and were checked not to allow for palindrome formation. Such design allowed to have every position on both strands of the DNA target (highlighted in blue in Table 2.1) to be methylated on either cytosine or adenine. In this design, the only adenine present in the PAM sequence was also methylated. These five target sequences were positioned in the middle of the 154 bp linear DNA target, which had PCR primer binding sequences integrated in them (Table 2.1). The sequences were synthesized by Metabion.

**Table 2.1 Linear DNAs and their PCR primer sequences**

PAM is red; target sequences are blue, and PCR primers are green. The sliding repeating motifs are underlined in single and double lines.

| Name | Sequence (5' to 3') |
|---|---|
| Target M1 | AGTCAAGTAACAACCGCGAAACCGCCGAATAACAGAGTAAGTCTCCGGTGG<br>TGTTCAGATTTTACGTAACGAATCGAATCGTAATGGTGATCAGCATGTCCCG<br>GATCGCGTCGTTATAGTGTGCGAGGCGTTCTTGGAGAGCAACTGCATAAGG |
| Target M2 | AGTCAAGTAACAACCGCGAAACCGCCGAATAACAGAGTAAGTCTCCGGTGG<br>TGTTCAGATTTTAGTAACGAATCGAATCGTAACTGGTGATCAGCATGTCCCG<br>GATCGCGTCGTTATAGTGTGCGAGGCGTTCTTGGAGAGCAACTGCATAAGG |
| Target M3 | AGTCAAGTAACAACCGCGAAACCGCCGAATAACAGAGTAAGTCTCCGGTGG<br>TGTTCAGATTTTATAACGAATCGAATCGTAACGTGGTGATCAGCATGTCCCG<br>GATCGCGTCGTTATAGTGTGCGAGGCGTTCTTGGAGAGCAACTGCATAAGG |
| Target M4 | AGTCAAGTAACAACCGCGAAACCGCCGAATAACAGAGTAAGTCTCCGGTGG<br>TGTTCAGATTTTAAACGAATCGAATCGTAACGATGGTGATCAGCATGTCCCG<br>GATCGCGTCGTTATAGTGTGCGAGGCGTTCTTGGAGAGCAACTGCATAAGG |
| Target M5 | AGTCAAGTAACAACCGCGAAACCGCCGAATAACAGAGTAAGTCTCCGGTGG<br>TGTTCAGATTTTAACGAATCGAATCGTAACGAATGGTGATCAGCATGTCCCG<br>GATCGCGTCGTTATAGTGTGCGAGGCGTTCTTGGAGAGCAACTGCATAAGG |
| Forward primer | AGTCAAGTAACAACCGCGAAACCGCCGAATAACAGAGT |
| Reverse primer | CCTTATGCAGTTGCTCTCCAAGAACGCCTCGCACACT |

2.2 sgRNAs

Each sgRNA was named after its perfectly matching target. All sgRNAs used in the study (their sequences are listed in Table 2.2) were synthesized by Synthego and diluted in 90% of RNase-free water and 10% TE buffer (both supplied by the manufacturer).

**Table 2.2 sgRNA sequences**

| Name | Sequence (5' to 3') |
|---|---|
| sgRNA_M1 | CGUAACGAAUCGAAUCGUAA |
| sgRNA_M2 | GUAACGAAUCGAAUCGUAAC |
| sgRNA_M3 | UAACGAAUCGAAUCGUAACG |
| sgRNA_M4 | AACGAAUCGAAUCGUAACGA |
| sgRNA_M5 | ACGAAUCGAAUCGUAACGAA |
| sgRNA_D | GUGAUAAGUGGAAUGCCAUG |

2.3 Linear DNA preparation for methylation experiments

The linear targets (from Table 2.1) were amplified with Q5 DNA polymerase (NEB, M0491L) in a 25 cycle PCR without GC enhancer, otherwise based on the manufacturer's protocol. After the correct PCR products were verified by 2% agarose (Thermo Scientific, R0492) gel electrophoresis in 1X TBE buffer (90mM Tris-base (Fisher Scientific, 10667243), 90mM boric acid (Carl Roth, 5935.2), 2mM EDTA (Carl Roth, 8040.2), pH=8 regulated with NaOH (Honeywell, 38215)), a GeneJET column purification kit (Thermo Scientific, K0702) was used as described by the manufacturer to clean each of the targets, which were eluted in Milli-Q water instead of the supplied elution buffer. 1µg of each clean PCR product was methylated on cytosine by M.SssI CpG methyltransferase (NEB, M0226L) or on adenine by EcoGII methyltransferase (NEB, M0603S) based on the manufacturer's protocol for both enzymes. The resulting methylated targets were cleaned from the reaction components with the GeneJET column purification kit and eluted in Milli-Q water.

2.4 Cas9 cleavage experiments

All separate components of the cleavage reaction (DNA, sgRNA, Cas9 (NEB, M0386M)) were diluted in 1X r3.1 buffer (NEB, M0386M; 1X buffer consists of: 100 mM NaCl, 50 mM Tris-HCl, 10

mM $MgCl_2$, 100 µg/ml Recombinant Albumin, pH 7.9 @ 25°C). The final ratio of sgRNA:Cas9:DNA in the reaction was 30:10:1. The sgRNA and *S. pyogenes* Cas9 were incubated for 15 minutes in 22 °C to hybridize before the addition of DNA, and the cleavage assay was also performed at 22 °C. Once the DNA was added to the reaction, aliquots at different timepoints (listed in Table 2.3) were transferred to stop solution (final concentration: 60mM EDTA (Carl Roth, 8040.2), 1U Proteinase K (Thermo Scientific, EO0492)) and incubated at 37 °C for 30 minutes. The linear DNA cleavage reaction products at each timepoint were visualized on a 10% native polyacrylamide gel electrophoresis (native PAGE) in 1X TBE buffer and post-stained with SYBR Gold (Invitrogen, S11494) before visualization on Amersham Typhoon. The target sequence was excised from supercoiled DNA samples with LguI (Thermo Scientific, FD1934) and size-selected by magnetic bead purification (Beckman Coulter, A63881). DNA cleavage reaction products were time-barcoded (NEB, E7645L). Part of the reaction was visualized with capillary electrophoresis, while the remaining reaction products were sequenced.

**Table 2.3 Timepoints used in each part of this study**

| | Linear on-target | Linear off-target | Supercoiling assay |
|---|---|---|---|
| 0 | | | |
| 10 seconds | | | |
| 30 seconds | | | |
| 1 minute | | | |
| 3 minutes | | | |
| 10 minutes | | | |
| 30 minutes | | | |
| 100 minutes | | | |
| 300 minutes | | | |
| 1440 minutes | | | |

## 2.5 Target library design

To assess the Cas9's promiscuity levels to negatively supercoiled targets, a previously generated library (Jones et al., 2021) was used. All sequences and design of the targets can be found in an aforementioned paper; however, an example of a member composition is also provided here in Figure

2.1. Briefly, the target library has many target sequence variations to create different off-targets when compared to sgRNA. These variations include single or double insertions, deletions, substitutions, and mismatches (Figure 2.1). However, there are 146 perfectly matched targets, which serve as a control. Around each target there is a buffer region of DNA ($Constant_L$ and $Constant_R$ in Figure 2.1), which controls for the same length of DNA. The buffer is especially important in the case of insertions and deletions as it evens the overall target length out. Then, these regions are followed by left and right barcodes, which are unique to each target and are used to tell each end of the cleaved target apart during sequencing. The barcodes are followed by primer sequences, which are used for PCR amplification of the library. In addition to this, at the end of the primers there are cut sites, required to remove plasmid backbone before sequencing. The remaining parts – right and left adapters, time barcode – are added after Cas9 cleavage assay to provide a cleavage time resolution of the targets during sequencing.

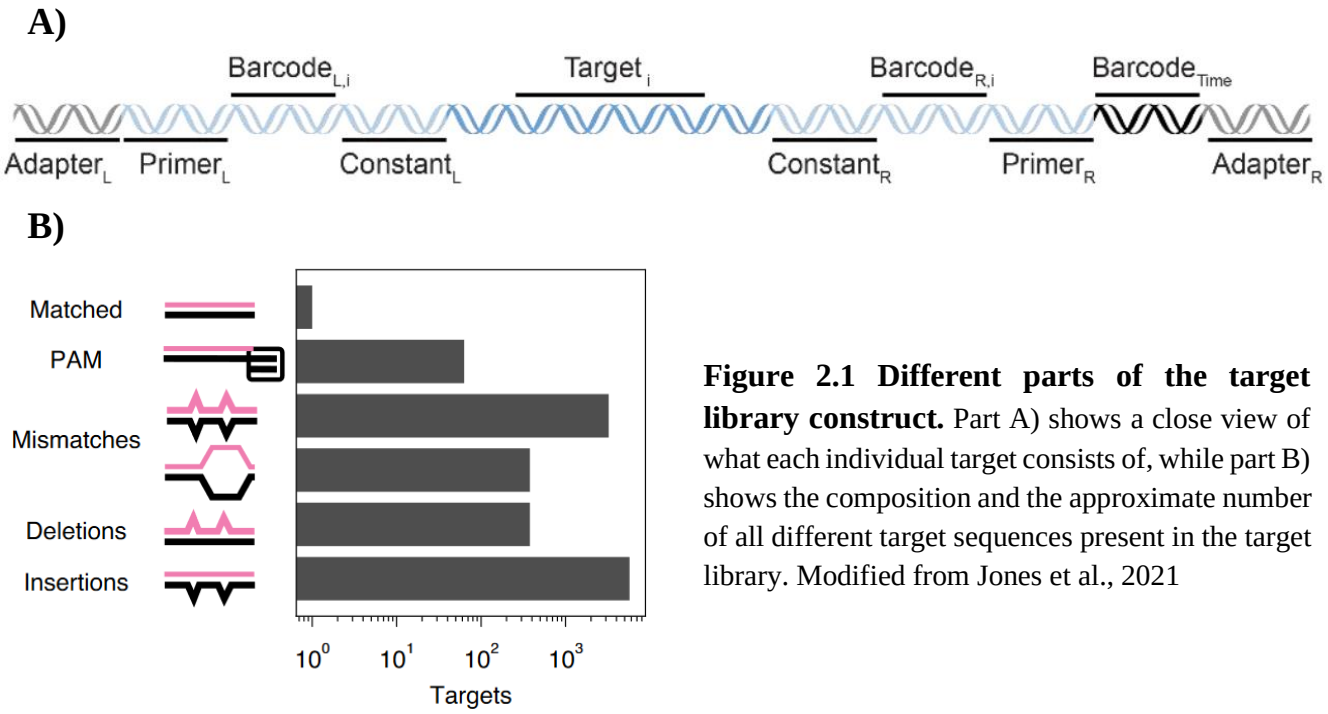


**Figure 2.1 Different parts of the target library construct.** Part A) shows a close view of what each individual target consists of, while part B) shows the composition and the approximate number of all different target sequences present in the target library. Modified from Jones et al., 2021

2.6 Construction of plasmid library

The plasmid was assembled from two parts – plasmid backbone and target library sequences (both supplied by Dr. Stephen Jones), and the assembled plasmid size was 1691bp. First, the plasmid backbone and target library were PCR amplified with Q5 DNA polymerase using primers from Table 2.4 in 25 cycles without GC enhancer, otherwise following the manufacturer's protocol. The correct

plasmid vector and target library PCR products were verified by 1% or 2% agarose gel electrophoresis in 1X TBE buffer, respectively. After the PCR products were column-purified, a HiFi master assembly mix (NEB, E2621S) was used to ligate the two pieces together. The ligation used 0.5 pmol of target library sequences and 0.1 pmol of plasmid backbone, and the rest of the reaction was carried out based on the manufacturer's protocol for using the assembly kit. The resulting ligation reaction was column-purified and eluted in Mili-Q water.

**Table 2.4 PCR primers for plasmid backbone and target library amplification**

| Name | Sequence (5' to 3') |
|---|---|
| Plasmid backbone, forward | ACTCTGTTATTCGGCGGTT |
| Plasmid backbone, reverse | AGTGTGCGAGGCGTTCTT |
| Target library, forward | AACCGCCGAATAACAGAGT |
| Target library, reverse | AAGAACGCCTCGCACACT |

2.7 Amplification of plasmid library

2μL of the column-purified plasmid library assembly reaction was transformed in 50μL of TOP10 electrocompetent *E. coli* cells (Thermo Scientific, C404050). After the electric pulse with a pulser (Bio-Rad; 1652660), the cells were resuspended in SOC media (Fisher Scientific, 11347689), pH=6.9, pre-warmed to 37°C, then recovered for 1 hour in 37°C with 250 rpm shaking. To test transformation efficiency, a 10 000-fold dilution was plated on a 100mm chloramphenicol (Carl Roth, 3886.2) petri dish (with a final antibiotic concentration of 25μg/mL), while the remaining cells in SOC media were further diluted to a final volume of 1.8 mL with LB broth (Carl Roth, X968.4). Then, the cells were divided equally among 3 150mm petri dishes, containing the same concentration of chloramphenicol. The smaller plate was incubated at 37°C while the big plates were incubated at 30°C for 22 hours. A single transformation was regarded as successful if the small plate had more than 90 colonies on it. As such competence was difficult to achieve during one transformation, two transformations were performed simultaneously. The combined transformation was regarded as successful if the combined number of colonies on both transformations' small plates was >90. All colonies from both transformations were combined into one sample during plasmid library purification.

2.8 Purification of plasmid library

After the transformation was successful, then each of the 150mm petri dishes were washed with LB broth 3 times, scraping the whole petri dish to resuspend all the cells on it. After each wash on each plate, the LB with the cells was collected in the same tube. Aliquots of 5mL of the resuspended cells were purified per single GeneJet's miniprep column (Thermo Scientific, K0702) based on the manufacturer's instructions, with two exceptions – purification was carried out on ice and eluted in Mili-Q water. The correct size of the plasmid was confirmed on 1% agarose gel electrophoresis in 1X TBE buffer. The bigger unwanted products were removed by gel-purifying only the correct-sized band with GeneJet's gel extraction kit (Thermo Scientific, K0692), following the manufacturer's instructions, but eluting in Mili-Q water instead of the elution buffer.

2.9 Plasmid library supercoiling production and density assessment

First, two different supercoiling levels of the plasmid library – fully relaxed and fully supercoiled – had to be obtained and verified, respectively. To fully relax purified plasmid, 6µg were exposed to Topoisomerase I from *E. coli* (NEB, M0301L) in its supplied reaction buffer (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/ml Recombinant Albumin, pH 7.9 @ 25°C. NEB, M0301L), performing the relaxation as per manufacturer's instructions. Relaxation was stopped with a stop buffer (final concentration: 0.25% SDS (Sigma-Aldrich, L5750) and 1mg/mL proteinase K) The reaction outcome was verified on 1% agarose electrophoresis in TBE buffer with SYBR Safe (Thermo Scientific, S33102) post-staining and ethanol precipitated with 0.066 µg/µL final concentration of glycogen (Thermo Scientific, R0561). To assess if the plasmid library remained supercoiled to the fullest after its isolation from *E. coli*, 0.2µg of untreated plasmid was exposed to gyrase in supercoiling buffer (35 mM Trizma-HCl (Sigma-Aldrich, T15760), pH 7.5 with NaOH, 25 mM KCl (Carl Roth, P017.1), 4 mM $MgCl_2$ (Fisher Scientific, AC223210010), 2 mM DTT (Fisher Scientific, 15875458), 1.8 mM spermidine (Fisher Scientific, 11449396), 1 mM ATP (Thermo Scientific, R0441), 6.25% (w/v) glycerol (Fisher Scientific, 12144481), 0.1 mg/mL BSA (Thermo Scientific, B14)), performing the reaction as recommended by the manufacturer and stopped with the stop buffer. The reaction products were visualized on 1% agarose in TBE buffer with SYBR Safe post-staining and compared to the untreated plasmid. An additional control of linearized plasmid for the assessment of relaxation and supercoiling reactions was also included. The linearization was performed with 0.5µg of plasmid using

LguI in its supplied reaction buffer (composition not specified; Thermo Scientific, FD1934) as per manufacturer's instructions.

The superhelical density assessment of plasmid library was performed following a 2D gel electrophoresis protocol that has been previously described (Martínez-García et al., 2018). Briefly, 161ng of both relaxed and supercoiled plasmid samples were mixed and loaded on 15cm×15cm 1% agarose gel in the first TBE buffer containing 0.6 μg/mL of chloroquine diphosphate. Then, the electric current was applied from top to bottom of the gel and ran at about 2-3 V/cm. The second direction run (left to right) was performed in TBE buffer containing 3 μg/mL of chloroquine diphosphate by manually twisting the gel 90° to the right and ran at the same voltage. After post-staining with SYBR Gold and visualising with Amersham Typhoon, the linking number distribution "steps" that appeared on the gel were calculated to find out the linking number difference (ΔLk). The linking number for relaxed DNA ($Lk_0$) was calculated using the equation (Liu et al., 2018):

$$Lk_0 = \frac{plasmid\ size\ in\ bp}{10.5 \frac{bp}{turn\ of\ DNA}}$$

These two numbers allowed to find out the superhelical density (σ) (Martínez-García et al., 2018):

$$\sigma = \frac{\Delta Lk}{Lk_0}$$

2.10 Data analysis

In order to obtain Cas9's cleavage rate from native PAGE gels, GelAnalyzer 19.1(Lazar & Lazar, n.d.) software was used to measure the DNA band intensity. When measuring the intensity, the background was removed by using a rolling ball option with a peak width tolerance in 20% of lane profile length. After this, the peaks were selected first based on the criteria that the peak threshold is 1 and the minimum height of the peak is 5. All peaks that did not satisfy these criteria but still corresponded to the bands of interest, were selected manually based on the position where the band was either visible or where it should be in the context of the surrounding bands, if invisible. The raw volume numbers obtained from this analysis were then normalized to be in a range from 0 to 1, using the formula below:

$$Normalized\ value = \frac{\frac{Uncut_t}{(Uncut_t + Cut_t)} - \frac{Uncut_f}{(Uncut_f + Cut_f)}}{\frac{Uncut_0}{(Uncut_0 + Cut_0)} - \frac{Uncut_f}{(Uncut_f + Cut_f)}}$$

In the formula, t stands for the raw volume at a timepoint, f is the raw volume for the final timepoint and 0 is the raw volume from timepoint 0.

Once the data for one cleavage assay was normalized, then it was fitted into an exponential decay function to give a cleavage rate in $s^{-1}$. Two models were used to obtain the best function fit to the dataset: $e^{-kt}$ or $e^{-kt}$ +c, where k is the cleavage rate, t is time, and c is horizontal asymptote, representing a fraction of DNA that never gets cut.

The data was analyzed using Past 4.03 (Hammer, 2001). First, the naked/methylated linear target cleavage data was checked for normal distribution with Shapiro-Wilk test and then analyzed with one-way ANOVA and Tukey's multiple comparison post-hoc test, or their non-parametric equivalents. The off-target data was first converted to log values and then analyzed. Data are present as means ± standard deviation.

# 3. RESULTS AND DISCUSSION

## 3.1 Methylation and on-target Cas9 cleavage

To evaluate if 5mC or 6mA DNA methylation alters the rates at which CRISPR-Cas9 cleaves on-target DNAs, we designed five targets (Figure 3.1). Across these targets, either cytosine or adenine could be enzymatically methylated (*i.e.*, 5mC and 6mA) at each position within a potential R-loop with Cas9's sgRNA. The PAM of all tested targets also had an adenine in it, which allowed us to test the impact of 6mA methylation on PAM identification. The results for on-target cleavage (Figure 3.2) indicated that cytosine methylation at any position does not alter Cas9 activity, as it did not change how fast any of the targets were cleaved when compared to their naked (no methyl group attached) counterparts. Investigating targets containing 6mA modifications revealed that targets M1 and M2 tended to be cleaved more slowly ($P>0.05$; Kruskal–Wallis one-way analysis of variance and Dunn's multiple comparisons) when their adenines were methylated. Meanwhile, target M4 had a significant reduction in its cleavage rate with 6mA methylation (Figure 3.2).

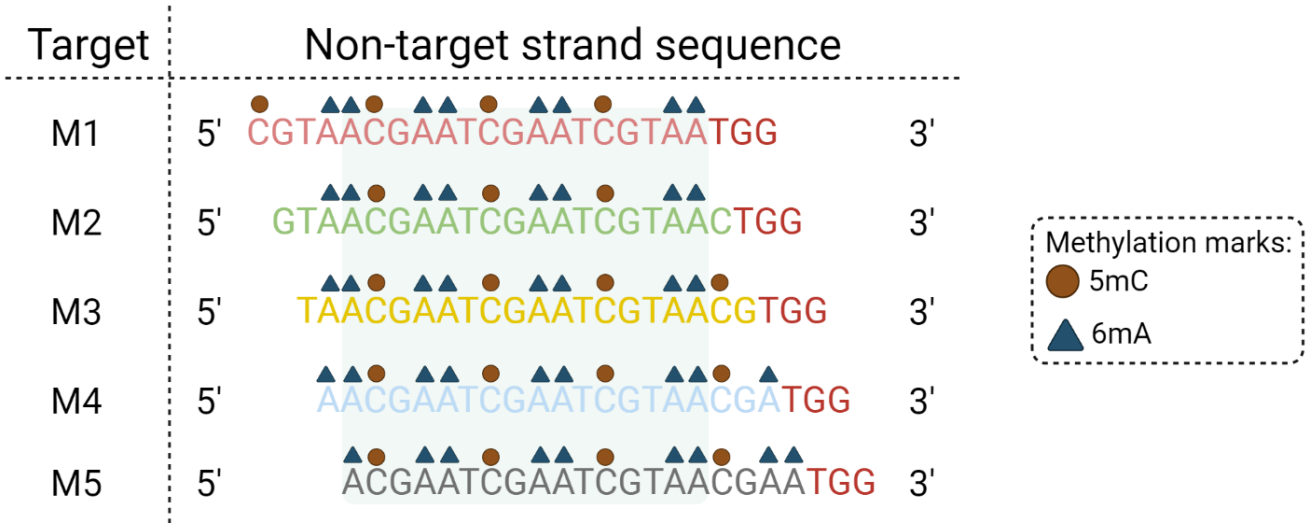


**Figure 3.1. Linear DNA target design.** Methylation targets 1-5 were designed to have a sliding sequence effect (see Methods section 2.1). The "travelling" sequence motif that is consistent throughout all five targets is highlighted by a mint box. Possible places for methylation on cytosine (5mC; circle) and on adenine (6mA; triangle) are showed only on the non-target strand. Note - the experiment never had both methylation types on the DNA at the same time. The figure was made by me with Biorender.com

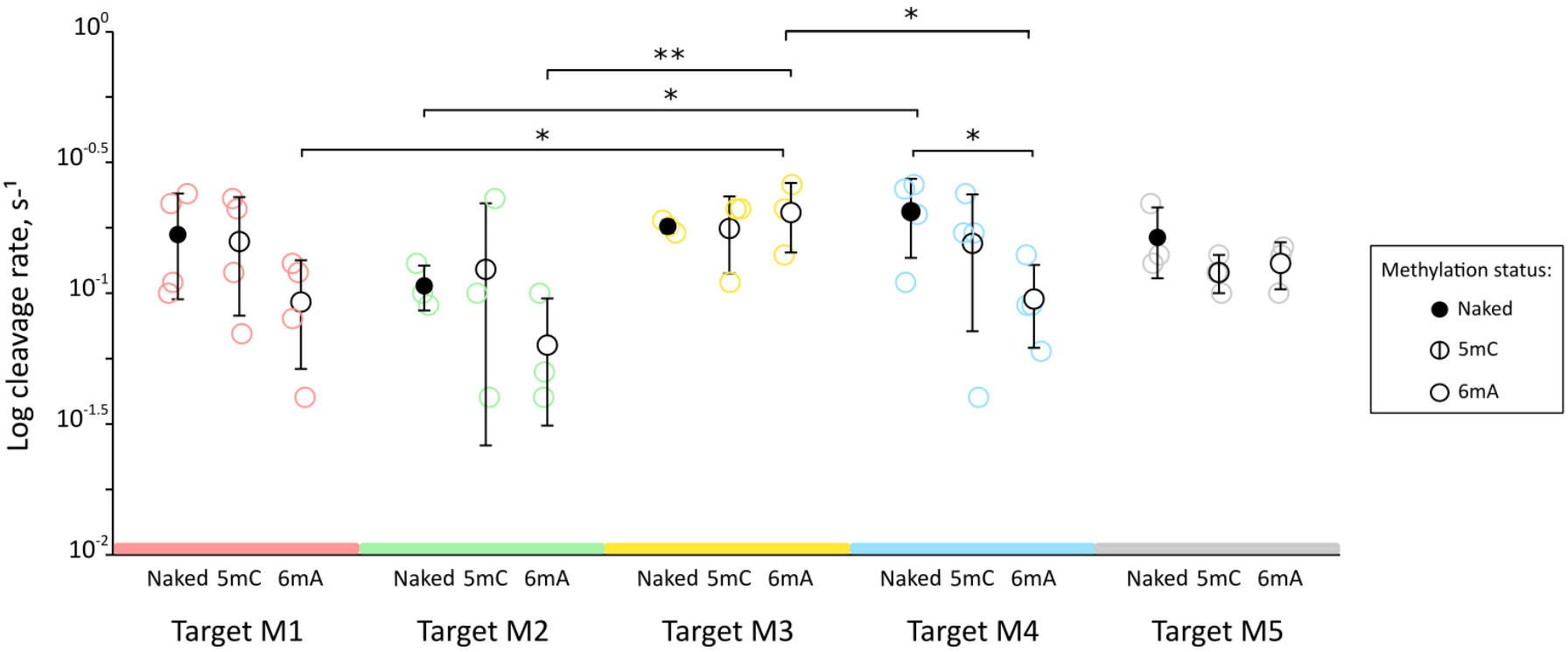


**Figure 3.2. On-target cleavage by *S. pyogenes* Cas9.** The targets for methylation study, named targets M1-M5, were targeted with their perfectly matching sgRNA. Cleavage reactions of naked (unmethylated) sequences and both methylated on cytosine (5mC) and adenine (6mA) sequences of each target were performed simultaneously and thus, each triplicate is time-matched. The reactions were carried out at 22°C and the $N \geq 3$ in each target. The data was tested for normal distribution using Shapiro-Wilk normality test and, as not all data was normally distributed, Kruskal–Wallis one-way analysis of variance and Dunn's multiple comparisons post hoc tests were applied. Data are present as mean ± SD, * is for $P<0.05$ and ** is for $P<0.01$

Why might some targets, such as M4, be cleaved differently when methylated, while others are not? Target M4 had an adenine as the first PAM-proximal base in the non-target strand and as the 6th PAM-proximal base on the target strand (Figure 3.1). It could be hypothesized that the most PAM-proximal position of the methylated adenine has the biggest effect on Cas9-gRNA interactions with target DNA (Figure 3.3) due to the seed region importance on the R-loop formation (Sternberg et al., 2014). But then target M1, which had two PAM-adjacent adenines on the non-target strand and an adenine on the 3rd PAM-proximal position of the target strand should have exhibited the biggest reduction in degradation rates, which was not the case. In addition to this, it has been shown that Cas9 is not even very sensitive to mismatches located in the very first two nucleotides adjacent to the PAM (Fu et al., 2014; Zeng et al., 2018), suggesting that methylation marks should also be well tolerated as they do not abolish sequence specificity when base pairing (Jeltsch, 2002).

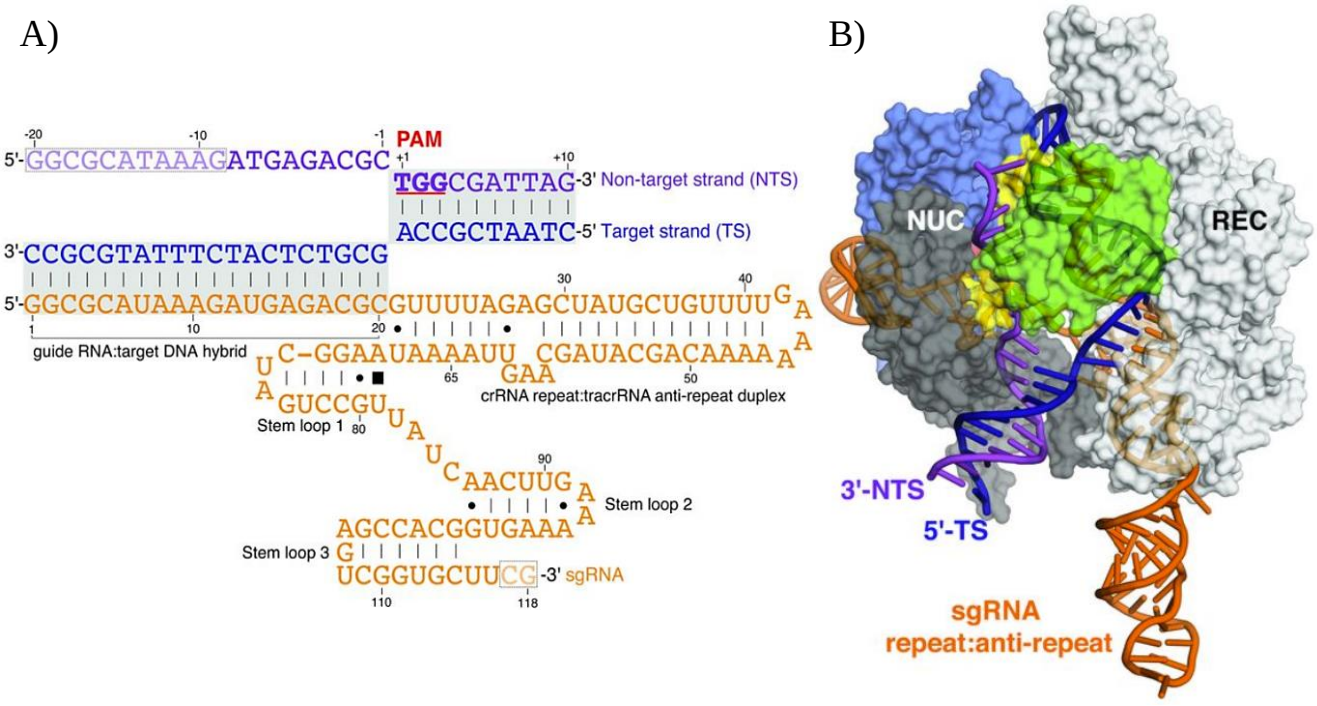


**Figure 3.3. *S. pyogenes* Cas9 R-loop.** Part A) shows a schematic representation of DNA and gRNA base-pairing. B) shows a crystal structure of Cas9 with its hybridized DNA target. NTS – non-target strand; TS – target strand; NUC – nuclease lobe; REC – helical recognition lobe. From Jiang et al., 2016.

To test if Cas9 detects methylation near DNA cleavage sites (the 3rd nucleotide on the target strand and the 3rd or the 4th on the non-target strand (Jiang and Doudna, 2017; Zeng et al., 2018)) we compared the sequences of our targets. Target M4 had an adenine on nucleotide 4 on the non-target strand and is cleaved slower, while, for example, M1 had an adenine on the 3rd nucleotide on the target

strand and its cleavage rate was unchanged when methylated. Both of these targets had an adenine on places that interact with different nuclease domains – could it be that one nuclease domain is more sensitive for methylation? A quick look at other sequences rejects this hypothesis. If having an adenine on the non-target strand is important (as is in M4), then M2 and M3 should have been cleaved slower when adenine-methylated, but it was not the case (Figure 3.2). Since the total amount of adenines on both DNA strands did not vary across targets (12), this also cannot explain why M4 is the only target to show a significant difference in its cleavage rate upon 6mA methylation. Thus, my data does not support a role for 6mA methylation proximity to the PAM, cleavage site, or in total as an explanation for why Cas9 may cleave such targets more slowly with this type of DNA methylation.

To investigate if target methylation also impacts how different sequences are interrogated, all five naked and methylated targets were compared to each other (Figure 3.2). Naked targets M2 and M4 showed a significant difference in their degradation rates ($P<0.05$; Kruskal–Wallis one-way analysis of variance and Dunn's multiple comparisons), but not if they were cytosine-methylated. Investigation of other targets did not show any differences. However, it is possible that this was due to insufficient statistical power, stemming from low replicate numbers ($N = 3$ or 4). If methylation status is important for on-target editing, incomplete methylation could explain these results. Methylation status was not explicitly checked – a decision we made since no consistent significant differences in cleavage rates were observed in my pilot experiments.

Cas9, when paired with different sgRNAs, has different on-target cleavage efficiency depending on the target sequence. If there is an adenine as the very first nucleotide adjacent to the PAM, this reduces the cleavage rate the most, while a cytosine here has an opposite effect (Gagnon et al., 2014; Liu et al., 2016). This contrasts with my results for targets M2 and M4, where the faster cleaved target M4 had an adenine, and target M2 had a cytosine (Figures 3.1 and 3.2). However, the first PAM-proximal nucleotide is not the only one dictating the cleavage rate – other nucleotides also have an effect. Liu and colleagues illustrated that the most favorable nucleotides within the seed region of non-target strand were adenine and cytosine (Liu et al., 2016). The sequences used in this study (Figure 3.1) were adenine and cytosine rich. The literature suggests that there should have been a difference in cleavage rates between targets M2 and M4 due to certain nucleotides in certain positions. However, they indicate a different direction in which each of the target cleavage rates should have changed (Gagnon et al., 2014; Liu et al., 2016). Thus, seed region nucleotides (including the most PAM-proximal one) are not sufficient to determine

cleavage rates. Perhaps it is not just the presence of these nucleotides in the seed, but instead their order? As it is known that the R-loop formation is the rate limiting step when it comes to cleavage by Cas9 (Gong et al., 2018), it would be beneficial to test with a bigger library if different order of nucleotides affect the R-loop formation kinetics.

Adenine-methylated target M3 was cleaved at a significantly faster rate than all other adenine-methylated targets, except for M5 (Figure 3.2). Apart from target M3, the other four adenine-methylated sequences displayed a very similar cleavage rate when compared to one another. While this could reflect incomplete methylation reactions or underpowered statistics, target sequence analysis shows that target M3 had fewer adenines immediately adjacent to its PAM than the other targets (Figure 3.1). It could be hypothesized that more adenines with methyl groups are proximal to the PAM sequence could reduce the cleavage rate, but in such a case, target M1 should have shown the slowest cleavage kinetics, which was not true. And to look at the other way around – the fewer adenines there are in the seed region, the more unaffected the cleavage rate is. This could be the case, as target M3 had a sequence with adenines positioned the furthest away from the PAM sequence. Yet, target M1 had the most PAM-proximal adenines, suggesting that the difference between their cleavage rates should have been bigger.

In respect to cleavage positions, M2 had methylated adenines next to the cleavage site on both target and non-target strands, which was also the case with M3. If this mattered, then both M2 and M3 would have the biggest cleavage rate reductions of all five targets, and M5 the lowest. Since only M4 showed a reduction in cleavage rate with methylated adenines ($P<0.05$; Kruskal–Wallis one-way analysis of variance and Dunn's multiple comparisons), not M3 or M2 when compared to their naked counterparts, this is unlikely.

The currently used design of target sequences had an adenine in their PAMs (Figure 3.1) and allowed us to see if placing a methyl group in it changes its cleavage kinetics. The results in Figure 3.2 show that the reduction in the rate of cleavage of adenine-methylated targets versus naked counterparts was insignificant ($P>0.05$; Kruskal–Wallis one-way analysis of variance and Dunn's multiple comparisons) for four out of five targets. Thus, methylating the adenine within the PAM sequence, which in our targets was 5'-TGG-3', does not interfere with *S. pyogenes* Cas9 ability to degrade its targets. Unfortunately, it was not possible to place a methyl group on the cytosines in the PAM with the methyltransferase used in this study.

Thus far, the only known methylation-sensitive Cas9 is from the thermophilic bacterium *Acidothermus cellulolyticus*. Its PAM sequence is 5'-NNNCC-3'. Only the first cytosine (4$^{th}$ nucleotide if reading the PAM from 5' to 3') induces exceptional Cas9's sensitivity for methylation, as modifying C4 abolished scission completely. This sensitivity is most likely achieved through target strand C4 interaction with Glu1044, non-target strand G4 interaction with Arg1088, and G5 with Arg1091 of the PAM-interacting domain (PID). This is in contrast to *S.pyogenes* Cas9, which recognises the PAM through interactions with guanines on the non-target strand (Das et al., 2020) (Figure 3.4). Knowing that sensitivity to methylation most likely requires the PID to interrogate both strands of DNA, it becomes unsurprising why *S. pyogenes* Cas9 did not respond to methylation on the adenine in the PAM tested here. However, it would be interesting to engineer an *S. pyogenes* Cas9 variant that would respond to methylation of its targets. This could be used as a quick means to test for the presence of methylation marks, by simply performing a digestion assay instead of bisulfite sequencing. Another potential application is the direct targeting of cancer cells, as they are known to lose their methylation status (Jung et al., 2020). This might allow the engineered Cas9 to cleave a target in cancer cells that would be otherwise inaccessible in healthy cells.

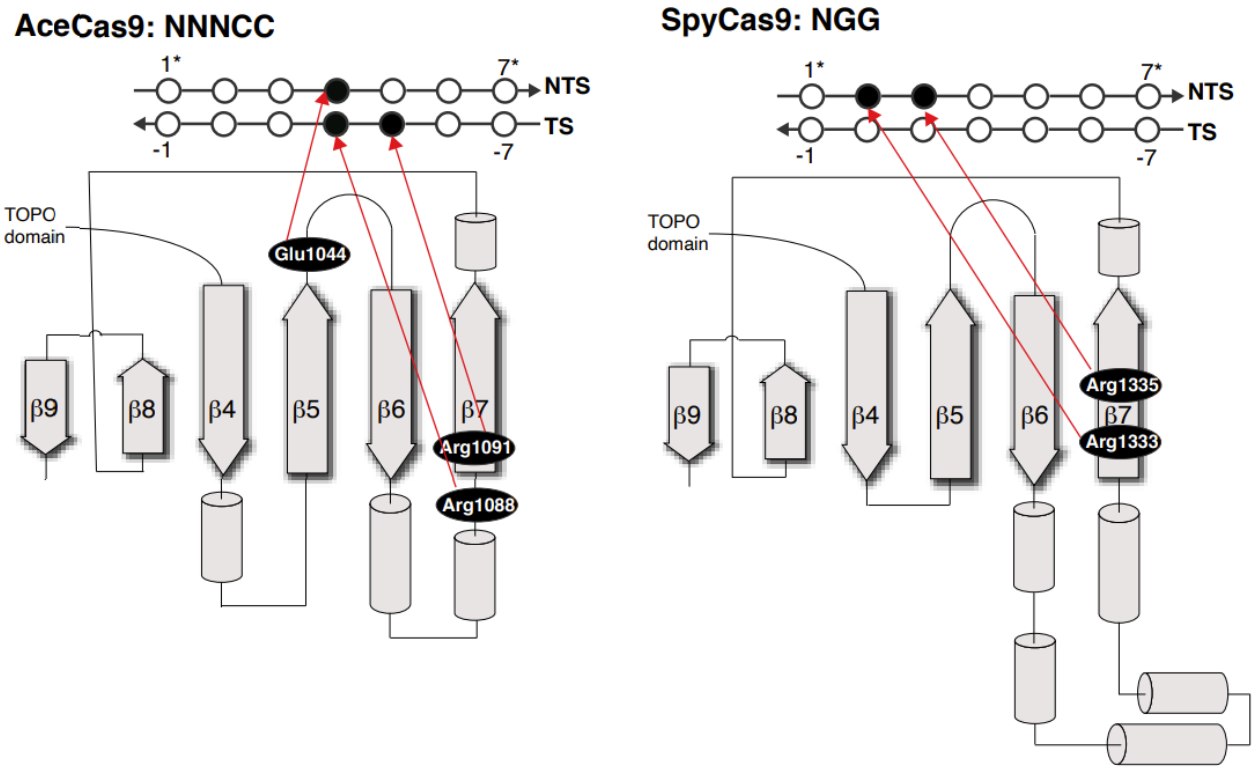


**Figure 3.4. PAM recognition by Cas9 from *A. cellulolyticus* (left) and *S. pyogenes* (right).** Modified from Das et al., 2020.

The investigation of on-target cleavage and the effect methylation has on it indicated that cytosine methylation does not affect *S. pyogenes* Cas9 cleavage kinetics. Adenine methylation, on the other hand, impacted the cleavage of one (target M4) out five targets investigated in this study. Analysing its sequence and the position of methyl group on target M4 failed to indicate a reason why this particular target was different from the other ones, thus more replicates and/or bigger library is needed to investigate this difference further. Due to the target design the PAM was also methylated, which, unlike for *A. cellulolyticus* Cas9 (Das et al., 2020), did not affect how sequences were interrogated. Finally, as it was reported previously (Gagnon et al., 2014; Jones et al., 2021; Liu et al., 2016), we observed varying cleavage rates for different, chemically unmodified on-targets.

3.2 Methylation and off-target cleavage

To understand how DNA methylation may impact Cas9 specificity, off-target cleavage was investigated for two out of five targets. As it was observed that the biggest difference between on-target cleavage rates was between targets M2 and M4 (Figure 3.2), it was decided to use the same targets to test off-target cleavage changes. Due to each target (and its related sgRNA) having a single base pair offset from our other targets (Figure 3.1), the most easily tested off-targets were those containing either a sgRNA-relative insertion or deletion in the target DNA. The off-targets investigated in this section were interrogated from two possible perspectives – target DNA and sgRNA. In the first case, two targets – M2 and M4 – were used with Cas9 paired to the sgRNAs for neighboring targets (sgRNA M1, M3, and M5). In the second case, Cas9 paired to sgRNA M3 was used with two neighboring DNA targets (M2 and M4) (Figure 3.5). Such combinations allowed us to examine cleavage of off-targets with a 1bp insertion or deletion relative to sgRNA M3 (inducing a DNA bulge or sgRNA bulge, respectively). Such formation of bulges is essential to execute Cas9's activity, as without them there would be far too many mismatches between the target and sgRNA (Dagdas et al., 2017; Sternberg et al., 2015).

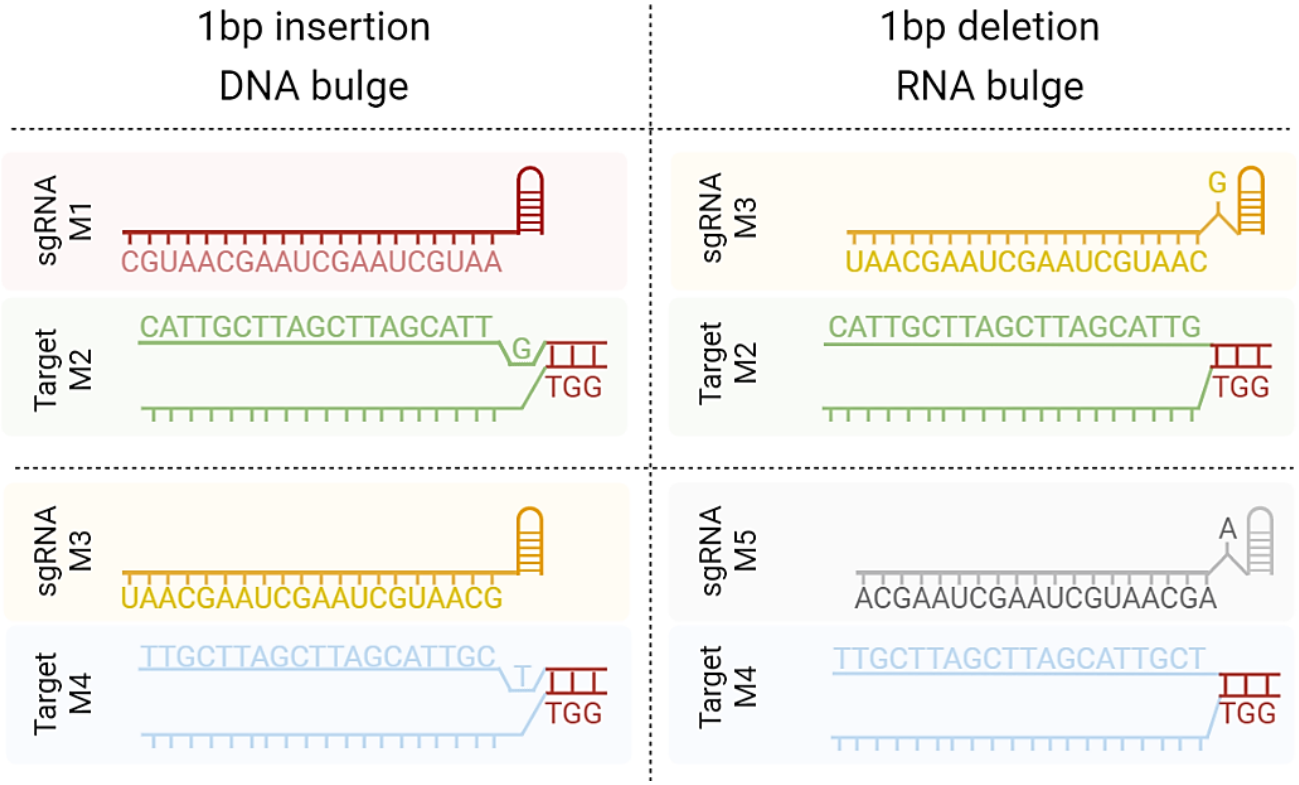


**Figure 3.5. Off-targets tested in this study.** Due to the sliding sequence design (Figure 3.1), targets M2 and M4 could have an insertion or deletion off-targets, forming DNA or RNA bulge, respectively. Made by me with Biorender.com

First, we wanted to investigate if methylation on cytosine or adenine has any effects on off-target cleavage. The results we obtained from the experiment (Figure 3.6), strongly suggested that methylation does not interfere with either DNA or RNA bulge-forming off-target cleavage (Figure 3.6, $P>0.05$; one-way ANOVA and Tukey's multiple comparisons post hoc). Interestingly, on-target analysis indicated that target M4 was cleaved slower with 6mA chemical modification (Figure 3.2), but this was not observed with off-target cleavage (Figure 3.6). It could be possible that this lack of difference between methylated and naked M4 counterparts in the off-target cleavage rate was because the rate is already substantially reduced and further reductions are negligible. Regardless, higher replicate numbers for both on- and off-target analyses are needed to give a sounder answer if there is a presence of significant differences.

We next tested if different off-targets are cleaved at different rates. Our results indicated that only M4 showed a significant reduction in Cas9 cleavage for all three types of DNA when comparing DNA and RNA bulge-forming off-targets (Figure 3.6). Target M2 did not show a difference in how fast it was degraded and, because of this, had a different cleavage profile from M4. This was rather expected as the on-target cleavage assay indicated a significant difference in their scission rates (Figure 3.2). It was also observed that DNA and gRNA bulge forming targets were cleaved at different rates. If looking from a DNA perspective, target M4 with sgRNA M5 (-1 bp) is cleaved slower than M4 with sgRNA M3 (+1 bp), if looking from sgRNA M3 perspective, then target M2 (-1 bp) has a significantly reduced scission rate than M4 (+1 bp). Therefore, my results suggest that RNA bulge-forming off-targets are much less tolerated than DNA bulge-forming ones.

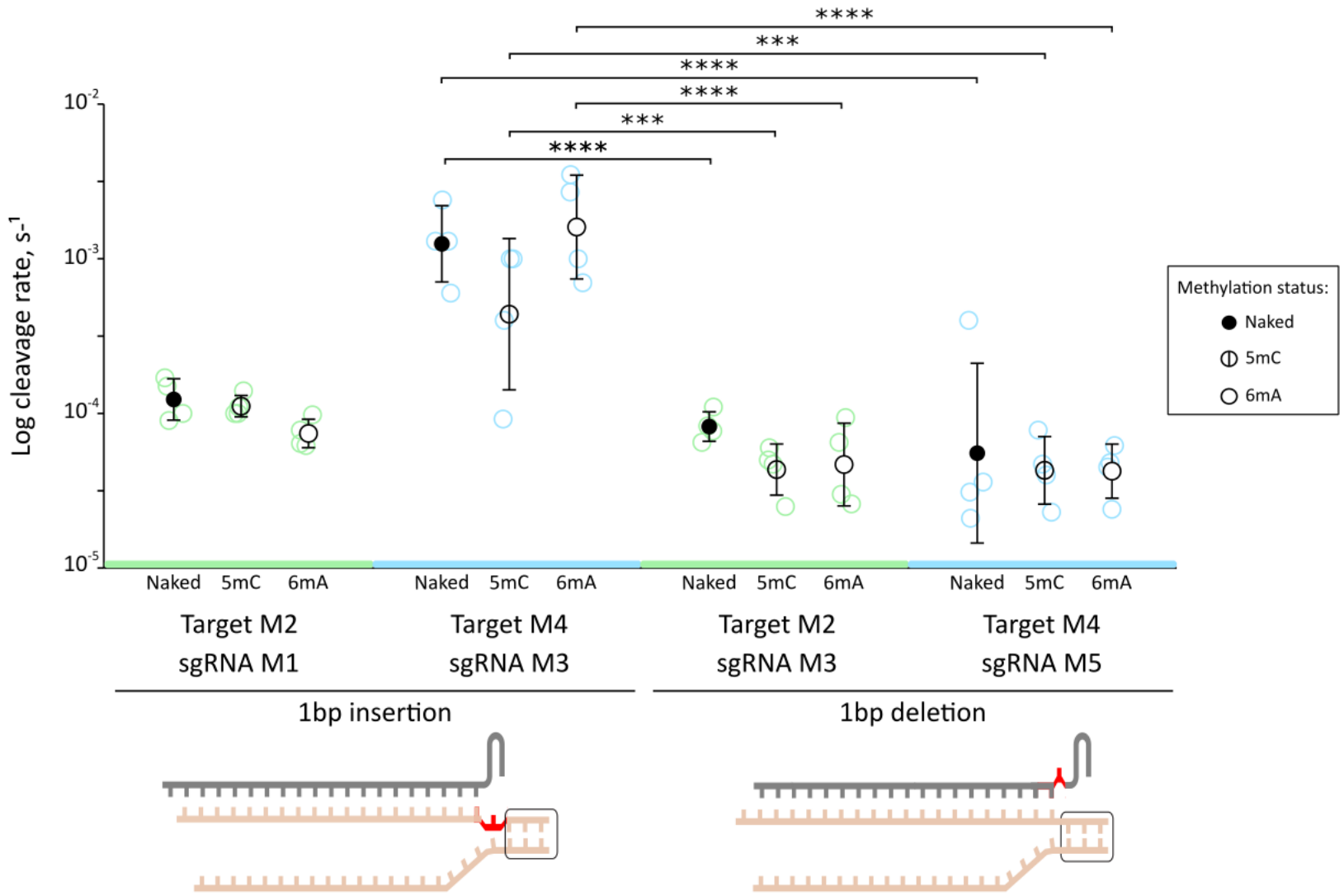


**Figure 3.6. Off-target cleavage by S. pyogenes Cas9.** Two off-targets were used with two sgRNAs to analyze the impact of RNA and DNA bulges on DNA cleavage by Cas9. Naked (unmethylated) sequences and both methylated on cytosine (5mC) and adenine (6mA) sequences of each target were performed simultaneously and thus, each triplicate is time-matched. The reactions were carried out at 22°C and the N=4 in each target. The data was converted to log and then tested for normal distribution using Shapiro-Wilk normality test followed by one-way ANOVA and Tukey's multiple comparisons post hoc tests. Data are present as mean ± SD, *** is for P<0.001 and **** is for P<0.0001

Could it be that such cleavage rate differences relate to different binding profiles? Previous work assessed if such DNA and sgRNA bulges affected dCas9 binding to DNA targets. Even though the work did not compare the binding rates of methylated and unmethylated DNA (Boyle et al., 2021), this study suggested that Cas9 does not usually sense methylation when cleaving, making it likely that the binding rates are unaffected as well. Back in 2020, Boyle and colleagues revealed that when there is only one nucleotide insertion induced bulge on DNA, the binding rate did not exhibit a severe decrease when compared to how simple 1bp mismatches were bound. However, sgRNA bulges showed a greater impact on dCas9 binding, with as little as a 3 nucleotide bulge in the PAM-proximal area leading to binding events below detection threshold. Interestingly, Cas9 and its target association rate was only weakly correlated to the cleavage rate, and slower binding did not mean slower cleavage. This explained why the cleavage rates for off-targets (with either a 1bp sgRNA-relative insertion and deletion) reduced to a much greater extent than the observed binding rates (Boyle et al., 2021). Our experiments examined single bulges on DNA or sgRNA and their cleavage rates, which agreed with Cas9 cleavage rates reported before (Boyle et al., 2021).

It has been shown that the identity of the bulging nucleotide on a target or sgRNA also has an effect on cleavage and binding rates of Cas9. The insertion of a cytosine reduce binding and cleavage rates the most, with cleavage being reduced by more than 100-fold. However, the insertion of a guanine has a relatively mild effect on binding and cleavage rates (Jones et al., 2021). One of the targets investigated here, target M2, had a cytosine insertion that slowed cleavage for the naked target from 0.10 $s^{-1}$ ± 0.02 to 0.00013 $s^{-1}$ ± 0.00038. Target M4 had a bulging adenine in relation to sgRNA M3; this was previously shown to have the second lowest reduction in cleavage rates (Jones et al., 2021). Here, we observed a similar phenomenon, as M4 cleavage slowed from 0.2 $s^{-1}$ ± 0.07 to 0.0014 $s^{-1}$ ± 0.00074, which is 10 times faster than that of target M2 containing a cytosine bulge.

The study performed here looked at two out of five designed targets (Figure 3.1) and their 1 nucleotide DNA or sgRNA bulge-forming off-target cleavage rate. It was found that RNA bulges are more detrimental to scission rate, however, neither bulge-forming off-target cleavage rate was affected by cytosine or adenine methylation. It would be interesting to investigate the remaining targets and also to investigate bigger bulges between DNA and sgRNA, as the current sequence design allows to test up to 4 nucleotide DNA and sgRNA bulges. Such future work could also assess if methylation starts having an effect if there are larger sequence differences between DNA and sgRNA.

### 3.3 DNA supercoiling and Cas9 cleavage

The second part of this project investigated to what extent negative supercoiling allows cleavage of off-targets and how much the target sequence correlates to this effect. To test this, a previously designed linear target library with different on- and off-target sequences ((Jones et al., 2021); Figure 2.1) was moved to a plasmid setting. Once this library and a plasmid backbone were assembled (see Methods 2.6), the construct was transformed to *E. coli*. The transformation's success was assessed (Table 3.1) to ensure the diversity of the library was maintained despite plasmid-to-plasmid variation. A single transformation did not meet our efficiency threshold. Thus, two different transformations were combined to ensure proper representation of each of the library members (see Methods 2.7).

**Table 3.1. Transformation quality assessment**

| | Library size | Copies of each library member | Final amount |
|---|---|---|---|
| Required | 8684 | 100 | At least 3µg |
| Observed | | 139 (54 + 67) | 32µg |

To assess how supercoiling affects cleavage by Cas9, we required relaxed and supercoiled configurations of the library. The supercoiling and relaxation reactions were tested separately (Figure 3.7). The Topoisomerase I-based relaxation reaction was checked using gel electrophoresis (Figure 3.7A), which confirmed the activity of topoisomerase I. The DNA gyrase-based supercoiling reaction was performed on purified plasmid, which should already be fully supercoiled due to the activity of endogenous *E. coli* enzymes. As expected, a linearized plasmid control migrated the least distance during gel electrophoresis when compared to both untreated plasmid and gyrase-treated plasmid (Figure 3.7B).

It was essential that our enzyme-treated plasmids represent single configurations, yet we observed that their size was not uniform (expected length: 1691bp). As these additional plasmids did not change after treatment with gyrase (Figure 3.7B), we hypothesized that they could be plasmid catenanes (two separate plasmid rings linked together (Higgins and Vologodskii, 2015)) or multiple library members fused end-to-end during plasmid assembly (see Methods 2.6). Catenanes would not affect our results, as there would be two targets in two plasmids that could be either relaxed or supercoiled. However, any incorrectly assembled plasmid (containing several targets) would mean that cleavage of one target would

linearize the supercoiled plasmid and the remaining targets in the plasmid would no longer be cleaved in a supercoiled state. To ensure the removal of these incorrect plasmid configurations (~20% of the total), the correctly-migrating plasmid was excised and gel-purified before moving to the next part of the study. Thus far, I have a correct-sized, clean plasmid library in two configurations – fully relaxed and fully supercoiled.

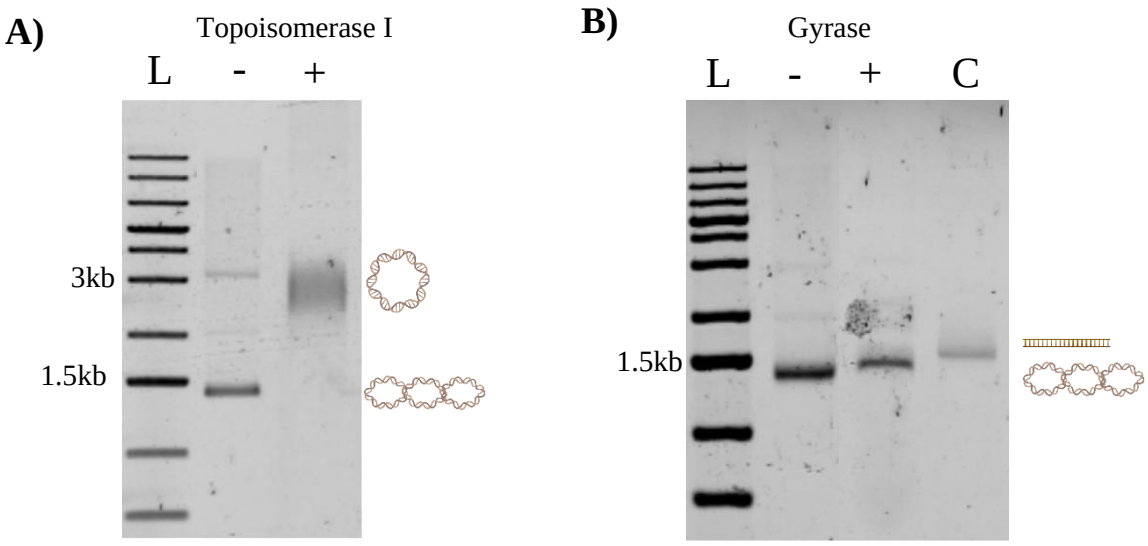


**Figure 3.7. Supercoiling and relaxing plasmid library.** The 1% agarose gel in the panel A) shows that topoisomerase I is capable of fully relaxing the plasmid. The same percentage gel in panel B) shows that the addition of a gyrase to purified sample does not result in a bigger supercoiling level of the plasmid. L is the ladder and C is the linearized plasmid control. Both gels were post-stained with SYBR Safe in order to prevent gel dye affecting plasmid migration.

## 3.4 Next steps and future work

To finish testing our hypothesis that negative supercoiling allows for off-target cleavage, we need to continue the started work. First, we need to test what is the level of superhelical density in our purified plasmid library. Next, we need to expose the supercoiled and relaxed plasmids for Cas9 degradation over time, prepare the samples for next-generation sequencing, sequence them, and then analyze the results.

Why do we care about supercoiling and off-target activity? Negative supercoiling affects how Cas9 cleaves on- and off-target DNAs, yet it is unknown if this effect is sequence-dependent or affects all DNAs similarly. Ivanov and colleagues back in 2020 showed that negative torque promoted the formation of an R-loop, while positive torque had an opposing effect. Therefore, the more negative the torque, the easier it is for Cas9 to start interrogating the sequences with more mismatches being tolerated. This effect has been illustrated by showing that Cas9 cleaved even off-target DNA if it was negatively supercoiled. However, linear substrates do not have the underwinding required to shift the energy landscape to the ready-to-cleave state, and did not allow off-target cleavage (Ivanov et al., 2020). However, the magnitude of this effect with target sequence analysis has not been done before, thus more data is needed. Given this information, we need to better understand how negative supercoiling affects DNA and sgRNA pairing when it comes to off-targets if their interactions become more energetically favorable. The second part of this project is focusing on investigating this effect in a greater detail. Once the data is analyzed, we could find some interesting things, such as maybe insertion or deletion-formed bulges are less supercoiling-state sensitive than simple mismatches.

The next step of our on-going project is the assessment of the superhelical density of our supercoiled plasmid library. Bacteria usually keep their plasmids at a superhelical density of -0.05 (Joyeux, 2015), however, we must be certain, as our future biophysical models will greatly depend on this. We showed that gyrase, also from *E. coli,* does not introduce additional, detectable supercoils (Figure 3.7B), likely because cellular gyrases already supercoiled our plasmid. Gyrases are energy-dependent enzymes, and limited by the energy released from ATP hydrolysis: *i.e.* the energy required to introduce a new supercoil is greater than what ATP hydrolysis provides (Martínez-García et al., 2018). We will confirm the level of supercoiling of our plasmid by performing 2-dimensional electrophoresis with chloroquine in the buffer. The method relies on different mobilities of relaxed and supercoiled plasmids in the presence of DNA intercalator to identify individual supercoiling states of a plasmid (Martínez-García et al., 2018).

Our final experiment, NucleaSeq (Jones et al., 2021), will capture the Cas9 cleavage kinetics of each target in our plasmid library. This includes the cleavage rate, the position of the cleavage, and any trimming happening after the cleavage. We will perform two biological replicates, which will be time-barcoded before sequencing. In the methylation part of the project the cleavage rate was obtained from running the cleavage reactions on a native polyacrylamide gel electrophoresis and quantifying band intensities (see Methods 2.4). This time, the cleavage rate will be obtained from sequenced data by comparing intact and cleaved sequences of any given target (for the types of different sequences see Methods 2.5) and how their proportion changed over time.

Once the cleavage rates will be obtained, our collaborators – Martin Depken and Hidde Offerhaus from TU-Delft – will generate a biophysical model of how on- and off-targets are interrogated when negatively supercoiled. The main limitation of previous models by one of the collaborators is the linear target DNAs on which the models were based (Eslami-Mossallam et al., 2022; Klein et al., 2018). They do not represent the state of the DNA inside the cells, as there it is typically negatively supercoiled (Corless and Gilbert, 2017; Kuzminov, 2014). Thus, using models based on linear DNA when trying to predict off-target activity in an *in vivo* setting might not give the most accurate data. Our work will eliminate this lack of information on the topic and will provide some very useful insights as it is already known that linear DNA has different off-target permissiveness than supercoiled DNA (Ivanov et al., 2020). It could be found that negative supercoiling tolerates more mismatches, DNA/sgRNA bulges, and still allows for cleavage, though the relaxed DNA does not. This could even affect the place of cleavage with certain off-targets, as the base pairing between the target and the guide would be affected by supercoiling and more favorable energetic landscape (Ivanov et al., 2020). There might even be some trimming events in response to some specific off-targets, which would be incorporated into the model to give the most thorough answer to how CRISPR-Cas9 responds to negatively supercoiled DNA.

## 3.5 Concluding remarks

The paper presented here illustrated that CRISPR-Cas9 is usually insensitive to DNA methylation. The on-target cleavage of cytosine methylated DNA is kinetically unaffected, which usually holds true for adenine-methylated DNA as well – it was significantly slowed down with only one out of five targets. Off-target cleavage rate does not change when comparing how DNA or sgRNA bulges are cleavage of naked and chemically modified DNA counterparts, however, DNA/sgRNA deletions are cleaved significantly slower than their corresponding insertions.

The paper also investigated if negative supercoiling affects how well off-targets are recognized by CRISPR-Cas9. Currently, a linear target library was successfully moved in to a plasmid and amplified. The library was either exposed to gyrase to introduce supercoils or fully relaxed by exposing to topoisomerase I. The future work will investigate the superhelical density level of the library and will expose both plasmid conformations to CRISPR-Cas9 digestion over time. Once the cleavage data will be sequenced and cleavage rates obtained, a biophysical model of how CRISPR-Cas9 interrogates different sequences of a negatively supercoiled substrate will be created. This will help us understand what target features affect nuclease activity and make its outcome more predictable for a safer application of CRISPR-Cas9 for gene editing.

# CONCLUSIONS

1. *S. pyogenes* Cas9 does not sense cytosine methylation when cleaving its on-targets.
2. *S. pyogenes* Cas9 only occasionally senses methylation on adenine in its on-targets.
3. Methylation status on either cytosine or adenine is not sensed by *S. pyogenes* Cas9 for off-targets that induce a single base RNA or DNA bulge.
4. RNA bulge forming off-targets are more detrimental for *S. pyogenes* Cas9 cleavage that DNA bulge forming off-targets.

# PARTICIPATION IN CONFERENCES

Ieva Jaskovikaitė, Stephen Knox Jones Jr. “The role of DNA topology in CRISPR-Cas9 nuclease specificity”. The Coins 2023, Vilnius, Lithuania. 2023-04-28

VILNIUS UNIVERSITY
LIFE SCIENCES CENTER

Ieva Jaskovikaitė
Master Thesis

# THE ROLE OF DNA TOPOLOGY IN CRISPR-CAS9 NUCLEASE SPECIFICITY

## SUMMARY

In bacteria, the clustered regularly interspaced short palindromic repeats (CRISPR) system with its associated protein 9 (CRISPR-Cas9) can recognize genetic material of invading bacteriophages and neutralize it. The CRISPR-Cas9 system is guided to its target with the help of guide RNA (gRNA), which makes it easily reprogrammable to a new target with just another gRNA and thus appealing for eukaryotic genome editing. However, a bacterial nuclease encounters different DNA topology when applied in a eukaryotic environment. The differences include different methylation patterns, as bacteria have cytosine and adenine methylation while eukaryotes only have cytosine methylation. Additionally, eukaryotic environment allows short-lived positive supercoils in addition to usual negative supercoils as opposed to bacteria. It has been shown that negative supercoiling is more off-target permissive (Ivanov et al., 2020), yet the magnitude of this effect is unknown. Therefore, the aim of the study is to evaluate if the topology (methylation and supercoiling) of on-target and off-target DNAs impacts their recognition and cleavage by CRISPR-Cas9 from *Streptococcus pyogenes.*

First, the study analyzed if cytosine or adenine methylation affected on- and off-target cleavage rates compared to unmethylated linear target DNAs. This work showed that CRISPR-Cas9 does not typically sense methylation on either nucleotide in its on-targets, with the exception of one out of five targets tested. Off-target cleavage rates are unaffected when methylated, with gRNA bulge forming off-targets having slower cleavage rate than DNA bulge forming off-targets.

It was also analyzed if negatively supercoiling DNA promotes CRISPR-Cas9 off-target cleavage when compared to relaxed DNA. Linear DNA target library was successfully inserted into a plasmid, amplified, then fully relaxed and supercoiled. Future work will describe cleavage rates, sites, and trimming for different targets and will produce a biophysical model of CRISPR-Cas9 cleavage specificity. The data will expand our knowledge on nuclease's specificity and help to choose the most specifically targeted places within the genome.

ACKNOWLEDGEMENTS

The second part of the project is supported by a collaboration with Prof. Martin Depken and Hidde Offerhaus from TU-Delft. They were an important and much valued part when it came to trouble-shooting and planning of supercoiling experiments. Once the data will be collected from the “wet-lab” experiments, they will be enormously contributing to the project by creating the biophysical model of Cas9s specificity.

I would like to thank Dr. Mindaugas Zaremba, for offering his very helpful hand to me when I was stuck with my experiments. Also, a big thank you for my fellow Jones!Lab members who always offered their emotional support and very good advice.

The biggest thank you of all goes to Dr. Stephen Jones for being the most patient, supportive, and resourceful mentor. I am forever grateful for these two amazing years and all the knowledge I have acquired.

REFERENCE LIST


1. Adhikari, S., Curtis, P.D., 2016. DNA methyltransferases and epigenetic regulation in bacteria. FEMS Microbiol. Rev. 40, 575–591. https://doi.org/10.1093/femsre/fuw023
2. Adli, M., 2018. The CRISPR tool kit for genome editing and beyond. Nat. Commun. 9, 1911. https://doi.org/10.1038/s41467-018-04252-2
3. 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, 1709–1712. https://doi.org/10.1126/science.1138140
4. Barrangou, R., Horvath, P., 2017. A decade of discovery: CRISPR functions and applications. Nat. Microbiol. 2, 1–9. https://doi.org/10.1038/nmicrobiol.2017.92
5. Beaulaurier, J., Schadt, E.E., Fang, G., 2019. Deciphering bacterial epigenomes using modern sequencing technologies. Nat. Rev. Genet. 20, 157–172. https://doi.org/10.1038/s41576-018-0081-3
6. Bendich, A.J., Drlica, K., 2000. Prokaryotic and eukaryotic chromosomes: what's the difference? BioEssays 22, 481–486. https://doi.org/10.1002/(SICI)1521-1878(200005)22:5<481::AID-BIES10>3.0.CO;2-T
7. Bikard, D., Marraffini, L.A., 2012. Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. Curr. Opin. Immunol., Innate immunity / Antigen processing 24, 15–20. https://doi.org/10.1016/j.coi.2011.10.005
8. Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiol. Read. Engl. 151, 2551–2561. https://doi.org/10.1099/mic.0.28048-0
9. Boyle, E.A., Becker, W.R., Bai, H.B., Chen, J.S., Doudna, J.A., Greenleaf, W.J., 2021. Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. Sci. Adv. 7, eabe5496. https://doi.org/10.1126/sciadv.abe5496
10. Broughton, J.P., Deng, X., Yu, G., Fasching, C.L., Servellita, V., Singh, J., Miao, X., Streithorst, J.A., Granados, A., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., Hsu, E., Gu, W., Miller, S., Pan, C.-Y., Guevara, H., Wadford, D.A., Chen, J.S., Chiu, C.Y., 2020. CRISPR–Cas12-based detection of SARS-CoV-2. Nat. Biotechnol. 1–5. https://doi.org/10.1038/s41587-020-0513-4
11. Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., van der Oost, J., 2008. Small CRISPR RNAs

Guide Antiviral Defense in Prokaryotes. Science 321, 960–964. https://doi.org/10.1126/science.1159689

12. Bryson, A.L., Hwang, Y., Sherrill-Mix, S., Wu, G.D., Lewis, J.D., Black, L., Clark, T.A., Bushman, F.D., 2015. Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. mBio 6, e00648-15. https://doi.org/10.1128/mBio.00648-15
13. Casadesús, J., 2016. Bacterial DNA Methylation and Methylomes, in: Jeltsch, A., Jurkowska, R.Z. (Eds.), DNA Methyltransferases - Role and Function, Advances in Experimental Medicine and Biology. Springer International Publishing, Cham, pp. 35–61. https://doi.org/10.1007/978-3-319-43624-1_3
14. Chedin, F., Benham, C.J., 2020. Emerging roles for R-loop structures in the management of topological stress. J. Biol. Chem. 295, 4684–4695. https://doi.org/10.1074/jbc.REV119.006364
15. Chen, X., Rinsma, M., Janssen, J.M., Liu, J., Maggio, I., Gonçalves, M.A.F.V., 2016. Probing the impact of chromatin conformation on genome editing tools. Nucleic Acids Res. 44, 6482–6492. https://doi.org/10.1093/nar/gkw524
16. Cofsky, J.C., Soczek, K.M., Knott, G.J., Nogales, E., Doudna, J.A., 2022. CRISPR–Cas9 bends and twists DNA to read its sequence. Nat. Struct. Mol. Biol. 29, 395–402. https://doi.org/10.1038/s41594-022-00756-0
17. Corless, S., Gilbert, N., 2017. Investigating DNA supercoiling in eukaryotic genomes. Brief. Funct. Genomics 16, 379–389. https://doi.org/10.1093/bfgp/elx007
18. Corless, S., Gilbert, N., 2016. Effects of DNA supercoiling on chromatin architecture. Biophys. Rev. 8, 245–258. https://doi.org/10.1007/s12551-016-0210-1
19. 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 Synth. Biol. 6, 428–438. https://doi.org/10.1021/acssynbio.5b00299
20. Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A., Yildiz, A., 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci. Adv. 3, eaao0027. https://doi.org/10.1126/sciadv.aao0027
21. 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. Nat. Commun. 11, 6346. https://doi.org/10.1038/s41467-020-20204-1

22. de Mendoza, A., Lister, R., Bogdanovic, O., 2020. Evolution of DNA Methylome Diversity in Eukaryotes. J. Mol. Biol., Reading DNA Modifications 432, 1687–1705. https://doi.org/10.1016/j.jmb.2019.11.003
23. 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, 602–607. https://doi.org/10.1038/nature09886
24. Dhar, G.A., Saha, S., Mitra, P., Nag Chaudhuri, R., 2021. DNA methylation and regulation of gene expression: Guardian of our health. The Nucleus 64, 259–270. https://doi.org/10.1007/s13237-021-00367-y
25. Doudna, J.A., Charpentier, E., 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096. https://doi.org/10.1126/science.1258096
26. Eslami-Mossallam, B., Klein, M., Smagt, C.V.D., Sanden, K.V.D., Jones, S.K., Hawkins, J.A., Finkelstein, I.J., Depken, M., 2022. A kinetic model predicts SpCas9 activity, improves off-target classification, and reveals the physical basis of targeting fidelity. Nat. Commun. 13, 1367. https://doi.org/10.1038/s41467-022-28994-2
27. Fu, B.X.H., Hansen, L.L., Artiles, K.L., Nonet, M.L., Fire, A.Z., 2014. Landscape of target:guide homology effects on Cas9-mediated cleavage. Nucleic Acids Res. 42, 13778–13787. https://doi.org/10.1093/nar/gku1102
28. Fujita, T., Yuno, M., Fujii, H., 2016. Allele-specific locus binding and genome editing by CRISPR at the p16INK4a locus. Sci. Rep. 6, 30485. https://doi.org/10.1038/srep30485
29. Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., Schier, A.F., 2014. Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. PLoS ONE 9, e98186. https://doi.org/10.1371/journal.pone.0098186
30. Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71. https://doi.org/10.1038/nature09523
31. Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V., 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. 109, E2579–E2586. https://doi.org/10.1073/pnas.1208507109

32. Gong, S., Yu, H.H., Johnson, K.A., Taylor, D.W., 2018. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. Cell Rep. 22, 359–371. https://doi.org/10.1016/j.celrep.2017.12.041
33. Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J., Verdine, V., Donghia, N., Daringer, N.M., Freije, C.A., Myhrvold, C., Bhattacharyya, R.P., Livny, J., Regev, A., Koonin, E.V., Hung, D.T., Sabeti, P.C., Collins, J.J., Zhang, F., 2017. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438–442. https://doi.org/10.1126/science.aam9321
34. Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6
35. Hatfull, G.F., Hendrix, R.W., 2011. Bacteriophages and their Genomes. Curr. Opin. Virol. 1, 298–303. https://doi.org/10.1016/j.coviro.2011.06.009
36. Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., Marraffini, L.A., 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature 519, 199–202. https://doi.org/10.1038/nature14245
37. Higgins, N.P., Vologodskii, A.V., 2015. Topological Behavior of Plasmid DNA. Microbiol. Spectr. 3, 3.2.17. https://doi.org/10.1128/microbiolspec.PLAS-0036-2014
38. Hinz, J.M., Laughery, M.F., Wyrick, J.J., 2015. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. Biochemistry 54, 7063–7066. https://doi.org/10.1021/acs.biochem.5b01108
39. Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R., Weissman, J.S., 2016. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. eLife 5, e12677. https://doi.org/10.7554/eLife.12677
40. 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. Nat. Biotechnol. 31, 827–832. https://doi.org/10.1038/nbt.2647
41. Hud, N.V., 1995. Double-stranded DNA organization in bacteriophage heads: an alternative toroid-based model. Biophys. J. 69, 1355–1362. https://doi.org/10.1016/S0006-3495(95)80002-0

42. Isaac, R.S., Jiang, F., Doudna, J.A., Lim, W.A., Narlikar, G.J., Almeida, R., 2016. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. eLife 5, e13450. https://doi.org/10.7554/eLife.13450
43. Ishino, Y., Krupovic, M., Forterre, P., 2018. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. J. Bacteriol. 200, e00580-17. https://doi.org/10.1128/JB.00580-17
44. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. 169, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987
45. Ivanov, I.E., Wright, A.V., Cofsky, J.C., Aris, K.D.P., Doudna, J.A., Bryant, Z., 2020. Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1913445117
46. Janik, E., Niemcewicz, M., Ceremuga, M., Krzowski, L., Saluk-Bijak, J., Bijak, M., 2020. Various Aspects of a Gene Editing System—CRISPR–Cas9. Int. J. Mol. Sci. 21, 9604. https://doi.org/10.3390/ijms21249604
47. Jansen, Ruud., Embden, Jan.D.A. van, Gaastra, Wim., Schouls, Leo.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x
48. Jeltsch, A., 2002. Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. ChemBioChem 3, 274–293. https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S
49. Jiang, F., Doudna, J.A., 2017. CRISPR–Cas9 Structures and Mechanisms. Annu. Rev. Biophys. 46, 505–529. https://doi.org/10.1146/annurev-biophys-062215-010822
50. Jiang, W., Chang, J., Jakana, J., Weigele, P., King, J., Chiu, W., 2006. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature 439, 612–616. https://doi.org/10.1038/nature04487
51. 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, 816–821. https://doi.org/10.1126/science.1225829

52. Jones, S.K., Hawkins, J.A., Johnson, N.V., Jung, C., Hu, K., Rybarski, J.R., Chen, J.S., Doudna, J.A., Press, W.H., Finkelstein, I.J., 2021. Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. Nat. Biotechnol. 39, 84–93. https://doi.org/10.1038/s41587-020-0646-5
53. Joyeux, M., 2015. Compaction of bacterial genomic DNA: clarifying the concepts. J. Phys. Condens. Matter 27, 383001. https://doi.org/10.1088/0953-8984/27/38/383001
54. Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F., Goel, A., 2020. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat. Rev. Gastroenterol. Hepatol. 17, 111–130. https://doi.org/10.1038/s41575-019-0230-y
55. Kallimasioti-Pazi, E.M., Chathoth, K.T., Taylor, G.C., Meynert, A., Ballinger, T., Kelder, M.J.E., Lalevée, S., Sanli, I., Feil, R., Wood, A.J., 2018. Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. PLOS Biol. 16, e2005595. https://doi.org/10.1371/journal.pbio.2005595
56. Kasman, L.M., Porter, L.D., 2023. Bacteriophages, in: StatPearls. StatPearls Publishing, Treasure Island (FL).
57. Klein, M., Eslami-Mossallam, B., Arroyo, D.G., Depken, M., 2018. Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. Cell Rep. 22, 1413–1423. https://doi.org/10.1016/j.celrep.2018.01.045
58. Kordyś, M., Sen, R., Warkocki, Z., 2022. Applications of the versatile CRISPR-Cas13 RNA targeting system. WIREs RNA 13, e1694. https://doi.org/10.1002/wrna.1694
59. Kouzarides, T., 2002. Histone methylation in transcriptional control. Curr. Opin. Genet. Dev. 12, 198–209. https://doi.org/10.1016/S0959-437X(02)00287-3
60. Kuzminov, A., 2014. The Precarious Prokaryotic Chromosome. J. Bacteriol. 196, 1793–1806. https://doi.org/10.1128/JB.00022-14
61. Lehman, I.R., Pratt, E.A., 1960. On the Structure of the Glucosylated Hydroxymethylcytosine Nucleotides of Coliphages T2, T4, and T6. J. Biol. Chem. 235, 3254–3259. https://doi.org/10.1016/S0021-9258(20)81347-7
62. Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., Bao, G., 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 42, 7473–7485. https://doi.org/10.1093/nar/gku402

63. Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., Wang, M., Zhang, X., Wang, Y., 2017. The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. Cell 170, 714-726.e10. https://doi.org/10.1016/j.cell.2017.06.050
64. Liu, M.-S., Gong, S., Yu, H.-H., Jung, K., Johnson, K.A., Taylor, D.W., 2020. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. Nat. Commun. 11, 3576. https://doi.org/10.1038/s41467-020-17411-1
65. Liu, X., Homma, A., Sayadi, J., Yang, S., Ohashi, J., Takumi, T., 2016. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. Sci. Rep. 6, 19675. https://doi.org/10.1038/srep19675
66. Liu, Y., Hua, Z.-C., Leng, F., 2018. DNA supercoiling measurement in bacteria. Methods Mol. Biol. Clifton NJ 1703, 63–73. https://doi.org/10.1007/978-1-4939-7459-7_4
67. Lowary, P.T., Widom, J., 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. 276, 19–42. https://doi.org/10.1006/jmbi.1997.1494
68. Luijsterburg, M.S., Noom, M.C., Wuite, G.J.L., Dame, R.Th., 2006. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. J. Struct. Biol. 156, 262–272. https://doi.org/10.1016/j.jsb.2006.05.006
69. Maeshima, K., Tamura, S., Hansen, J.C., Itoh, Y., 2020. Fluid-like chromatin: Toward understanding the real chromatin organization present in the cell. Curr. Opin. Cell Biol., Cell Nucleus 64, 77–89. https://doi.org/10.1016/j.ceb.2020.02.016
70. Marenduzzo, D., Orlandini, E., Stasiak, A., Sumners, D.W., Tubiana, L., Micheletti, C., 2009. DNA–DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. Proc. Natl. Acad. Sci. 106, 22269–22274. https://doi.org/10.1073/pnas.0907524106
71. Marks, T., Sharp, R., 2000. Bacteriophages and biotechnology: a review. J. Chem. Technol. Biotechnol. 75, 6–17. https://doi.org/10.1002/(SICI)1097-4660(200001)75:1<6::AID-JCTB157>3.0.CO;2-A
72. Martínez-García, B., Valdés, A., Segura, J., Dyson, S., Díaz-Ingelmo, O., Roca, J., 2018. Electrophoretic Analysis of the DNA Supercoiling Activity of DNA Gyrase. Methods Mol. Biol. Clifton NJ 1805, 291–300. https://doi.org/10.1007/978-1-4939-8556-2_15

73. McGinn, J., Marraffini, L.A., 2016. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. Mol. Cell 64, 616–623. https://doi.org/10.1016/j.molcel.2016.08.038
74. Mojica, F. j. m., Ferrer, C., Juez, G., Rodríguez-Valera, F., 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. Mol. Microbiol. 17, 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x
75. 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. J. Mol. Evol. 60, 174–182. https://doi.org/10.1007/s00239-004-0046-3
76. Mojica, F.J.M., Juez, G., Rodriguez-Valera, F., 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol. Microbiol. 9, 613–621. https://doi.org/10.1111/j.1365-2958.1993.tb01721.x
77. Nakata, A., Amemura, M., Makino, K., 1989. Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. J. Bacteriol. 171, 3553–3556. https://doi.org/10.1128/jb.171.6.3553-3556.1989
78. Naureen, Z., Dautaj, A., Anpilogov, K., Camilleri, G., Dhuli, K., Tanzi, B., Maltese, P.E., Cristofoli, F., De Antoni, L., Beccari, T., Dundar, M., Bertelli, M., 2020. Bacteriophages presence in nature and their role in the natural selection of bacterial populations. Acta Bio Medica Atenei Parm. 91, e2020024. https://doi.org/10.23750/abm.v91i13-S.10819
79. Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D., Higgins, C.F., 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48, 297–310. https://doi.org/10.1016/0092-8674(87)90433-8
80. 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, 935–949. https://doi.org/10.1016/j.cell.2014.02.001
81. Parashar, N.C., Parashar, G., Nayyar, H., Sandhir, R., 2018. N6-adenine DNA methylation demystified in eukaryotic genome: From biology to pathology. Biochimie 144, 56–62. https://doi.org/10.1016/j.biochi.2017.10.014

82. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., Liu, D.R., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31, 839–843. https://doi.org/10.1038/nbt.2673
83. Pourcel, C., Salvignol, G., Vergnaud, G., 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653–663. https://doi.org/10.1099/mic.0.27437-0
84. Ren, R., Horton, J.R., Zhang, X., Blumenthal, R.M., Cheng, X., 2018. Detecting and interpreting DNA methylation marks. Curr. Opin. Struct. Biol., Protein–nucleic acid interactions • Catalysis and regulation 53, 88–99. https://doi.org/10.1016/j.sbi.2018.06.004
85. Schmitz, R.J., Lewis, Z.A., Goll, M.G., 2019. DNA Methylation: Shared and Divergent Features across Eukaryotes. Trends Genet. 35, 818–827. https://doi.org/10.1016/j.tig.2019.07.007
86. Schvartzman, J.B., Hernández, P., Krimer, D.B., Dorier, J., Stasiak, A., 2019. Closing the DNA replication cycle: from simple circular molecules to supercoiled and knotted DNA catenanes. Nucleic Acids Res. 47, 7182–7198. https://doi.org/10.1093/nar/gkz586
87. Seong, H.J., Han, S.-W., Sul, W.J., 2021. Prokaryotic DNA methylation and its functional roles. J. Microbiol. 59, 242–248. https://doi.org/10.1007/s12275-021-0674-y
88. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., Zhang, F., 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science 343, 84–87. https://doi.org/10.1126/science.1247005
89. Shi, D.-Q., Ali, I., Tang, J., Yang, W.-C., 2017. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. Front. Genet. 8.
90. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., Severinov, K., Zhang, F., Koonin, E.V., 2017. Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169–182. https://doi.org/10.1038/nrmicro.2016.184
91. Simon, J.A., Kingston, R.E., 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. 10, 697–708. https://doi.org/10.1038/nrm2763
92. Stern, M.J., Ames, G.F.-L., Smith, N.H., Clare Robinson, E., Higgins, C.F., 1984. Repetitive extragenic palindromic sequences: A major component of the bacterial genome. Cell 37, 1015–1026. https://doi.org/10.1016/0092-8674(84)90436-7

93. Sternberg, S.H., LaFrance, B., Kaplan, M., Doudna, J.A., 2015. Conformational control of DNA target cleavage by CRISPR–Cas9. Nature 527, 110–113. https://doi.org/10.1038/nature15544
94. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–67. https://doi.org/10.1038/nature13011
95. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., Aryee, M.J., Joung, J.K., 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197. https://doi.org/10.1038/nbt.3117
96. Vlot, M., Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., Dickman, M.J., Brouns, S.J.J., 2018. Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR–Cas effector complexes. Nucleic Acids Res. 46, 873–885. https://doi.org/10.1093/nar/gkx1264
97. Weigel, C., Seitz, H., 2006. Bacteriophage replication modules. FEMS Microbiol. Rev. 30, 321–381. https://doi.org/10.1111/j.1574-6976.2006.00015.x
98. Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., Wigley, D.B., 2019. Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. Mol. Cell 0. https://doi.org/10.1016/j.molcel.2019.04.020
99. Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K., Carroll, D., 2018. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proc. Natl. Acad. Sci. 115, 9351–9358. https://doi.org/10.1073/pnas.1810062115
100. Yaung, S.J., Esvelt, K.M., Church, G.M., 2014. CRISPR/Cas9-Mediated Phage Resistance Is Not Impeded by the DNA Modifications of Phage T4. PLOS ONE 9, e98811. https://doi.org/10.1371/journal.pone.0098811
101. Yuan, X., Huang, Z., Zhu, Z., Zhang, J., Wu, Q., Xue, L., Wang, J., Ding, Y., 2023. Recent advances in phage defense systems and potential overcoming strategies. Biotechnol. Adv. 65, 108152. https://doi.org/10.1016/j.biotechadv.2023.108152
102. Zeng, Y., Cui, Y., Zhang, Yong, Zhang, Yanruo, Liang, M., Chen, H., Lan, J., Song, G., Lou, J., 2018. The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. Nucleic Acids Res. 46, 350–361. https://doi.org/10.1093/nar/gkx1117

103. Zhang, R., Xu, W., Shao, S., Wang, Q., 2021. Gene Silencing Through CRISPR Interference in Bacteria: Current Advances and Future Prospects. Front. Microbiol. 12.