

The Biology of Native and Adapted CRISPR-Cas Systems



Jack D. Sanford^{1,2,3*}, John E. Weldon¹

Clustered regularly interspaced short palindromic repeats (CRISPR) systems have revolutionized the life sciences since their development as an experimental tool in 2012. Native CRISPR systems act in prokaryotes as an adaptive immune system against invading genetic elements, such as viral DNA. These systems recognize invading nucleic acids, insert segments of the sequence in the host genome, and use these sequences to recognize and destroy the viral element if the cell is invaded again. In recent years, proteins from CRISPR systems, particularly the Cas9 nuclease, have been repurposed for different applications, such as gene editing experiments, large scale genetic screens, and imaging of DNA elements. CRISPR systems have dramatically increased the ease and efficiency of genome engineering, and further investigation and development of these systems is likely to continue for years to come. This literature review was written to consolidate and analyze the available research into native type II CRISPR systems and explore the breadth of adapted CRISPR systems. The simplicity and versatility of CRISPR systems have made them far superior to previous genome engineering technologies. The generation of improved Cas9 nucleases, discovery of additional CRISPR systems, and characterization of Cas9 off-target cleavage will expand the use of CRISPR systems in the lab and increase their viability in the clinic. In the few years since they achieved prominence, CRISPR-Cas9 systems have spurred new lines of inquiry in the biological sciences and provided a robust new toolkit to researchers. It is important for both novice and advanced scientists to understand the origins, applications, and limitations of these CRISPR systems.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR), along with CRISPR-associated (Cas) genes, are a form of adaptive immune system that has been found in ~40% of bacteria and ~80% of archaea (Makarova et al., 2011). These systems rely on the formation of sequences known as spacers in the CRISPR region of the host genome. Spacer sequences are identical to sections of invading foreign nucleic acids, commonly from phages. These spacer regions are transcribed into noncoding CRISPR-RNA (crRNA), which act as guides to direct an effector nuclease to make targeted cuts in invading genetic material. Targeted cleav-

age of invading DNA prevents expression of viral elements, which prevents successful infection of the bacterium. The type II CRISPR system of *Streptococcus pyogenes* requires only one effector protein, Cas9, which can be targeted to make a double-stranded break in DNA at a specific nucleotide sequence (Jinek et al., 2012).

Modified CRISPR systems, the vast majority of which use the Cas9 protein, have become revolutionary tools for genetic modification for two main reasons: ease of use and high versatility. Previous methods to modify the genomes of organisms have also relied on the introduction of double-stranded breaks, but were difficult and expensive to design (Doudna & Charpentier, 2014). Examples of this include zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Doudna & Charpentier, 2014). CRISPR systems, however, require only the design of a guide RNA complementary to a target site. Recent developments have created numerous modified CRISPR systems, which use the targeted Cas9 protein for purposes beyond the standard double-stranded cleavage (Brocken et al., 2017; B. Chen et al., 2013; Cheng et al., 2013; Nishida et al., 2016; Qi et al., 2013). This review covers a brief history of CRISPR research, what is known about the biology of the native type II CRISPR system, and several of the numerous different CRISPR-based applications that have been developed in recent years. Adapted CRISPR systems have proven to be incredibly effective tools for biological and biomedical research due, in large part, to their versatility. Although Cas9 originally evolved to simply cleave invading viral elements in single-celled organisms, it has been used in adapted CRISPR systems to make targeted genetic and epigenetic alterations, image DNA elements, alter gene expression, and discover key genes in-

¹ Department of Biological Sciences, Jess and Mildred Fisher College of Science and Mathematics, Smith Hall, Room 312, Towson University, 8000 York Rd, Towson, MD 21252

² Curriculum in Genetics and Molecular Biology, University of North Carolina-Chapel Hill, 120 South Rd, Chapel Hill, NC 27599

³ University of North Carolina Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, 101 Manning Dr, Chapel Hill, NC 27599

*To whom correspondence should be addressed:
jsan4d@email.unc.edu



Except where otherwise noted, this work is licensed under <https://creativecommons.org/licenses/by/4.0>

doi:10.22186/jyi.35.5.81-91

volved in several processes. Additionally, through the study of native CRISPR systems, the discovery of the Cas13a protein in *Leptotrichia buccalis* has led to the development of a CRISPR-based system for the identification of specific DNA sequences in human tissue samples. Continued research into the use of catalytically deactivated Cas9 (dCas9) tethered to protein functional domains will continue to increase the number of possible functions for CRISPR systems. Further study of the diverse variety of CRISPR systems present in bacteria and archaea may lead to the discovery of new Cas proteins, each with their own functions, which will expand the versatility of adapted CRISPR systems.

A Brief History of CRISPR Research from Discovery to Application

The initial observation of a repeated 29 nucleotide segment in the genome of the bacteria *Escherichia coli* spent years in obscurity before a few enterprising scientists raised its profile enough to catch the attention of a larger field of researchers (Ishino et al., 1987). Subsequent advances in both the basic and applied scientific arenas came rapidly, stimulating more research. Several articles from different perspectives have been published regarding the history of CRISPR research (Doudna & Charpentier, 2014; Lander, 2015; Morange, 2015). This review is not meant to repeat their summaries but will highlight several key steps in the development of CRISPR as a tool.

Although it was not known as CRISPR at the time, the first CRISPR region was identified in 1987 by Atsuo Nakata and colleagues, who discovered strange repeats in the *iap* gene in *Escherichia coli* (Ishino et al., 1987). They found five adjacent nucleotide repeats, 29 nucleotides in length, interspaced with regions 32 nucleotide in length. As more genomic data became available, similar regions were identified in other bacteria as well as several archaea (Jansen, et al., 2002). Although it was an interesting observation, no function was assigned to the sequence. Research groups studying these sequences later agreed on the nomenclature CRISPR, an acronym for clustered regularly interspaced short palindromic repeats, to describe the sequence composition of the genomic region (Jansen et al., 2002). In 2002, researchers identified four CRISPR-associated (*cas*) genes, which they named *cas1-cas4* (Jansen et al., 2002). A study following up on this observation found additional *cas* genes in 27 organisms across two phyla of archaea and ten phyla of bacteria, expanding the number of known *cas* genes to 45 (Haft, Selengut, Mongodin, & Nelson, 2005).

A functional relationship for CRISPR/Cas was identified in 2005 by three groups, who each independently found that CRISPR spacers were derived from foreign genetic elements (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). The group of Elena Soria thus hypothesized that CRISPR systems may act to confer resistance to specific invading DNA elements, such as viruses (Mojica et al., 2005). To support this hypothesis, mutant bacterial strains were created with experimentally inserted CRISPR spacers corresponding to specific bacteriophages; these bacteria gained resistance to only those specific phages (Barrangou et al., 2007).

In 2009, it was discovered that CRISPR systems use a proto-

spacer-adjacent motif (PAM) to distinguish between self and non-self (Mojica et al., 2009). The protospacer is the region on the invading DNA corresponding to the targeting spacer. PAMs are short sequences directly adjacent to the protospacer on targeted DNA. Different CRISPR systems have different PAM sequence requirements; the *S. pyogenes* CRISPR/Cas9 system has an NGG PAM sequence requirement (Mojica et al., 2009). Another breakthrough occurred in 2010, when it was discovered that CRISPR systems function by creating double-stranded breaks in invading phage and plasmid DNA (Garneau et al., 2010).

A major step in the process of understanding a biological system is the ability to reconstitute it outside of its natural host. This was made possible by the discovery that the *S. thermophilus* CRISPR system only requires one effector protein: Cas9 (Sapranauskas et al., 2011). Researchers transferred the *S. thermophilus* CRISPR system, including the genes *cas9*, *cas1*, *cas2*, *csn2*, and a specialized CRISPR array onto a plasmid (Sapranauskas et al., 2011). *E. coli* transformed with this plasmid were resistant to genetic material which contained sequences corresponding to the CRISPR spacers (Sapranauskas et al., 2011). Of the protein-coding genes in the system, only deletion of the *cas9* gene eliminated resistance (Sapranauskas et al., 2011).

CRISPR development was taken one step further by recreating a functional CRISPR system *in vitro*, with only isolated proteins and nucleic acids (Jinek et al., 2012). A group led by Jennifer Doudna and Emmanuelle Charpentier found that *S. pyogenes* Cas9 nuclease and two noncoding RNA species (crRNA and tracrRNA) were the only required components for targeted endonuclease activity (Jinek et al., 2012). They also created a chimeric RNA, which combined the crRNA and tracrRNA, generating a two-component system for targeted genomic editing (Jinek et al., 2012). The system was subsequently engineered for human genome editing using an expression vector containing *S. pyogenes cas9* with a codon sequence optimized for mammalian cell expression and a nuclear localization signal (Cong et al., 2013). This, combined with a chimeric single guide RNA (sgRNA) for targeting, was used to engineer human and murine cell lines (Cong et al., 2013). This 2013 publication by a group led by Feng Zhang was a milestone in the CRISPR revolution and set off a rush to explore and adapt the system. The Addgene plasmid repository (www.addgene.org) has compiled a free resource on CRISPR that details its history and use as a genome-editing tool (Addgene, 2017).

Diversity of CRISPR-Cas Systems

Before 2011, the nomenclature for different CRISPR systems and Cas proteins was fragmented, confusing, and did not reflect the evolutionary relationships of CRISPR systems. In 2011, however, a review coauthored by many of the leaders in the CRISPR field established a new system for CRISPR nomenclature that is easier to use and more accurately accounts for the origins of the different CRISPR systems (Makarova et al., 2011; Makarova et al., 2015). The authors classified CRISPR systems as type I, type II, or type III based on their different effector mechanisms and nucleic acid targets. Type I and type III CRISPR systems have multi-subunit ef-

factor complexes. Type I systems use the hallmark Cas3 protein for target cleavage, while type III systems use either Csm or Cmr effector proteins, which target invading DNA or RNA, respectively. Type II CRISPR systems, which are the focus of applied CRISPR technologies and this review, have only one effector protein, Cas9. In these systems, Cas9 is activated by a tracrRNA, and targeted using a crRNA. Interaction with the tracrRNA induces an activating conformational change in Cas9, while the crRNA directs Cas9 to the targeted DNA sequence (Jinek et al., 2012, 2014; Nishimasu et al., 2014).

Native Type II CRISPR-Cas9 Function

CRISPR systems work in three steps that must occur in sequence: adaptation, expression, and interference (Figure 1) (van der Oost et al., 2014). During adaptation, the organism incorporates a section of the invading nucleic acid into its CRISPR array (van der Oost et al., 2014). Expression is when all the necessary components, including the crRNAs from the CRISPR array, are expressed and combined into an active nuclease complex (van der Oost et al., 2014). Interference occurs when the system targets and cleaves invading genetic material. Type II CRISPR systems are the simplest known variation of this process (van der Oost et al., 2014). The type II-A system of *S. thermophilus* involves only six components, Cas1, Cas2, Csn2, Cas9, tracrRNA, and the targeting crRNA (Heler et al., 2015). All elements except the crRNA are required for the adaptation step, while only the tracrRNA, crRNA, and Cas9 are required for the interference step (Jinek et al., 2012; Wei et al., 2015).

Adaptation

Adaptation is the key process that allows organisms to incorporate invading DNA into their genomes and, due to its complexity, is also the least understood process of the three. In *S. thermophilus* Cas1, Cas2, Csn2, Cas9, and the tracrRNA are all required for adaptation, but the Cas9 nuclease activity is not (Wei et al., 2015). Csn2 co-purifies with Cas1, Cas2, and Cas9, suggesting that these proteins form a stable complex (Heler et al., 2015). This complex is hypothesized to play a role in the adaptation process because of the known roles of Cas1 and Cas2 in spacer acquisition (Nuñez et al., 2014). In *S. pyogenes*, Csn2 forms a tetrameric diamond structure which exhibits the capacity to bind and travel along DNA (Arslan et al., 2013). A recent crystal structure revealed that two Cas1 proteins can bind to either side of the Csn2 tetramer, forming a hexamer (Ka et al., 2016). The positioning of Csn2 does not interfere with the *E. coli* Cas1-Cas2 complex crystal structure (Ka et al., 2016). This suggests that Csn2 may act as a scaffold on which Cas1 and Cas2 form a complex surrounding target DNA molecule. Cas9 is also known to interact with Cas1, Cas2, and Csn2, likely to grant PAM specificity to spacer acquisition, but the nature of this interaction has not been well described (Heler et al., 2015).

In the type I CRISPR system of *E. coli*, Cas1 has been shown to be a metal-dependent nuclease required for spacer acquisition (Nuñez et al., 2014). Interfering with the metal ion binding pocket of *S. pyogenes* Cas1 (SpCas1) prevents spacer acquisition, demonstrating that Cas1 is required for spacer acquisition in *S. pyogenes*

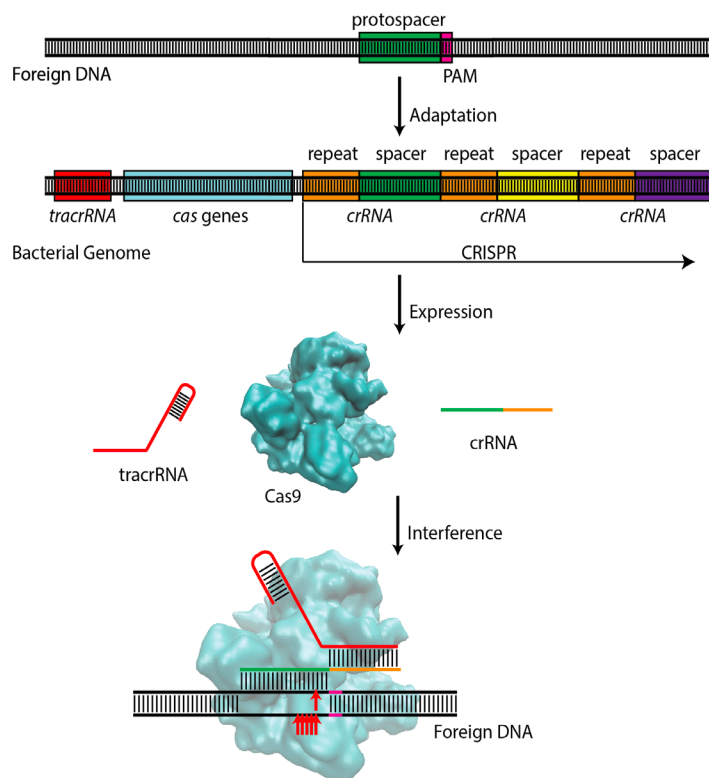


Figure 1. Native type II CRISPR system. The type II CRISPR system functions through three phases: adaptation, expression, and interference. During adaptation, a protospacer sequence (green) adjacent to a PAM site (pink) is recognized in foreign DNA and incorporated into the CRISPR array in the genome of the host. The CRISPR region consists of repeated sequence (orange) interspaced with spacer sequences (green, yellow, purple) identical to foreign genetic sequences. Each *crRNA* gene contains both spacer and repeat sequence. The CRISPR-associated (*cas*) genes (cyan) and *tracrRNA* gene (red) are located nearby. During the expression phase, these genes are transcribed, and the *cas* mRNA is subsequently translated. The crRNA, tracrRNA, and Cas9 protein form a complex that targets foreign DNA for cleavage during the interference phase. The crRNA forms base pairing interactions with complementary sequence in the foreign DNA and the tracrRNA. Cleavage (red arrows) occurs at a specific site 3 base pairs from the PAM sequence on the complementary strand, but is more variable on the non-complementary strand and can occur at multiple sites 3-8 base pairs from the PAM sequence.

as well (Heler et al., 2015). It is still unclear, however, whether SpCas1 acts as a nuclease like *E. coli* Cas1 does. A recent crystal structure of SpCas1 and SpCsn2 showed that the conserved amino acids responsible for metal ion binding were not in sufficient proximity to function together properly (Ka et al., 2016). In addition, metal ions were not co-purified with SpCas1 and SpCas1 did not exhibit nuclease activity, suggesting that Cas1 may not act as a nuclease in *S. pyogenes* (Ka et al., 2016). This result may be an artefact of the His6-MBP purification tag, or it may indicate that additional unknown cofactors are required for SpCas1 nuclease activity (Ka et al., 2016). Further study is necessary to resolve the

role of Cas1.

The role of Cas2 in type II CRISPR spacer acquisition is also not well understood. Cas2 is a metal- and pH-dependent endonuclease which has been shown to cleave both ssRNA and dsDNA (Nam et al., 2012). *Bacillus halodurans* (CRISPR type I subclass C) Cas2 forms a symmetric dimer to create a single binding pocket between the two proteins (Nam et al., 2012). It has been hypothesized that Cas1 and Cas2 work in concert to create dsDNA protospacers of the correct size because neither protein independently fragments DNA to the ~40bp size observed in CRISPR arrays (Nam et al., 2012). Contradictory results were found with *E. coli* (CRISPR type I subclass E), for which Cas2 enzymatic activity was not required for spacer acquisition (Nuñez et al., 2014). Instead, it has been shown that the addition of Cas2 leads to increased CRISPR locus binding of Cas1, which indicates that Cas2 may play a role in CRISPR site recognition (Nuñez et al., 2014). Due to these contradicting results, it remains unclear whether Cas2 enzymatic activity is important for spacer acquisition. It may be that Cas2 enzymatic activity is required for spacer acquisition in *B. halodurans*, but not *E. coli*.

The role of Cas9 in type II CRISPR adaptation appears to include PAM recognition. In one study, if the PAM recognition domain of Cas9 is removed, spacers without adjacent PAM sites were acquired (Heler et al., 2015). This suggests that Cas9's role in spacer acquisition is the identification of PAM sites on foreign DNA, which can later be targeted by the CRISPR system. It has also been shown that in the type II *S. thermophilus* CRISPR system, if the nuclease activity of Cas9 is deactivated, the majority of all new spacers found (96%) correspond to the bacterial genome (Wei et al., 2015). This suggests the possibility that CRISPR may be a high-risk, high-reward process in which the majority of newly created CRISPR spacers target the bacterial genome and result in self-cleavage and the death of the bacterium. The experiments performed by Wei et al. (2015) were conducted by transforming Cas gene plasmids into Cas deletion strains of *S. thermophilus*. So, it is unclear whether self-targeting CRISPR spacers would be added to the CRISPR array in the absence of plasmid transformation. They also suggest that another Cas9-dependent system could either prevent self-cleavage or remove self-targeting spacers.

After a spacer sequence is removed from the invading DNA element, it must be inserted into the bacterial genome. In *E. coli*, this insertion is catalyzed by the spacer integration complex, the crystal structure of which was recently solved by Jennifer Doudna and colleagues (Nuñez et al., 2015; Wright et al., 2017). This complex consists of two central Cas2 proteins which are flanked by two Cas1 dimers, creating a heterohexamer. The 33-nucleotide spacer DNA is held above the protein complex, which contacts the DNA on its opposite side. Integration is achieved by nucleophilic attack of the spacer DNA sequence by the protospacer DNA ends. Another protein, termed the integration host factor (IHF) aids in both target specificity and integration efficacy by sharply bending the DNA upstream of the integration site, which brings an upstream interaction site in contact with the integration complex

while creating a conformation which favors integration. Following integration, the viral DNA element can be expressed along with the interference machinery to prevent future viral infections.

Expression

After spacers are acquired, they must be expressed and processed. Recent research suggests that bacteria may be able to upregulate Cas gene expression in response to phage infection (Patterson et al., 2017). Some bacteria may be able to upregulate Cas gene expression in response to several infection-related stimuli, such as quorum-sensing signals from neighboring bacteria, disruptions in the bacterial membrane, and alterations in the levels of specific intracellular metabolites like cAMP (Patterson et al., 2017). Upregulation of Cas gene expression in response to viral infection has also been detected at the mRNA level in *Thermus thermophilus* (Agari et al., 2010) and *Sulfolobus solfataricus* (Fusco et al., 2015), and at the protein level in *Streptococcus thermophilus* (Young et al., 2012). It remains unclear, however, whether these modes of CRISPR activation are widespread or restricted to specific organisms.

Interestingly, these studies show regulation of Cas gene expression, while few studies show altered expression of crRNA or pre-crRNA expression (Patterson et al., 2017). This may suggest that CRISPR expression regulation focuses on modulating expression of effector protein levels rather than the expression of the targeting RNA components. In *S. pyogenes*, crRNA is constitutively transcribed as a ~511 nucleotide pre-crRNA which is then cleaved to form 39-42 nucleotide mature crRNA for targeting during the interference steps (Deltcheva et al., 2011). This processing step involves Cas9, RNase III, and tracrRNA (Deltcheva et al., 2011). The tracrRNA, another component of the CRISPR operon, pairs with the pre-crRNA and the two are co-processed by RNase III and Cas9 to form the correctly sized crRNA (Deltcheva et al., 2011).

Interference

The interference step in type II CRISPR systems requires Cas9 bound to a complex of the tracrRNA paired with a targeting crRNA (Jinek et al., 2012). *S. pyogenes* Cas9 is a multifunctional, 1368 amino acid protein (UniProtKB Q99ZW2) (Jinek et al., 2014; Nishimasu et al., 2014). Recent crystal structures of catalytically deactivated Cas9 molecules bound to nucleic acid substrates have given great insight into the exact function of Cas9 proteins (Jinek et al., 2014; Nishimasu et al., 2014). The protein has two lobes, a larger recognition (REC) lobe, and a smaller nuclease (NUC) lobe connected by a single bridge helix. The NUC lobe has two nuclease domains, an HNH and a RuvC domain. When not bound to a crRNA, these domains are located too far away from each other to act in concert (Jinek et al., 2014). However, after crRNA binds to Cas9, a drastic conformational change occurs which brings the two nuclease domains together. In this active conformation, the top and bottom lobes are separated by a shallow groove sufficiently wide to hold an RNA-DNA duplex (Jinek et al., 2014). The HNH and RuvC domains flank the groove, indicating that both could cleave target DNA simultaneously (Jinek et al., 2014). The groove is highly positively charged, further suggesting its role as a DNA-binding site (Jinek et al., 2014). Consistent with the crystal struc-

ture, mutated Cas9 variants have been used to show that the HNH domain cleaves the strand complementary to the crRNA while the RuvC domain cleaves the non-complementary strand (Gasiunas et al., 2012). In order to distinguish between invading DNA elements and the DNA located in the CRISPR array, Cas9 will only cleave DNA with an adjacent PAM sequence (Jinek et al., 2012).

This PAM specificity is likely achieved by residues in a topoisomerase homology domain (Jinek et al., 2014). In a crystal structure of crRNA and DNA bound to SpCas9, the PAM sequence is located at the junction of the two nuclease lobes, adjacent to Trp⁴⁷⁶ and Trp¹¹²⁶ (Jinek et al., 2014). These tryptophan residues are conserved among all Cas9 species that target an *NGG* PAM sequence, and are required for SpCas9 activity (Jinek et al., 2014). Cross-linking studies also show that these tryptophan-containing loops contact the PAM sequence in DNA-bound SpCas9 (Jinek et al., 2014). Researchers have suggested that these tryptophan residues might form base-stacking interactions allowing SpCas9 to identify the *NGG* PAM site, and could explain how Cas9 functions during spacer acquisition (Jinek et al., 2014). If Cas9 is not bound to a crRNA, the NUC domain would not be active, but the REC domain may still allow Cas9 to identify PAM sequences for the spacer acquisition complex. Binding crRNA also regulates Cas9 because it cannot cleave target DNA without being in complex with crRNA.

In the hypothetical model, Cas9 complexes with the tracrRNA cofactor and the targeting crRNA. Binding the tracrRNA-crRNA duplex leads to a conformational change in Cas9, which creates a binding groove in which Cas9 can interact with target DNA sequences. If Cas9 then encounters a sequence complementary to its crRNA and contains the requisite PAM sequence, nuclease domains then simultaneously cleave the target sequence. Mismatches at the 5' end of the guide sequence can be tolerated, but the final 8 base pairs at the 3' end of the guide must be an exact match. Mismatches in this so-called "seed sequence" greatly disrupt Cas9 cleavage activity (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). Cas9 cleaves both strands of the targeted DNA sequence 3 nucleotides upstream of the PAM site (Cong et al., 2013; Jinek et al., 2012). Further 3'-5' exonuclease activity trims an additional three to eight nucleotides off of the noncomplementary strand (Jinek et al., 2012). This creates a double-stranded break that can leave 5' overhangs on the invading DNA (Jinek et al., 2012).

Adapted CRISPR Systems

Cas9 is the sole effector protein in type II CRISPR systems (Makarova et al., 2015). This makes type II CRISPR systems ideal for genetic manipulation experiments. The most basic CRISPR systems require only Cas9 bound to a crRNA-tracrRNA duplex (Cong et al., 2013; Jinek et al., 2012). Researchers have also developed a chimeric RNA which combines the crRNA and the tracrRNA, making for a two-component targeted genome editing system (Jinek et al., 2012). The chimeric single guide RNA (sgRNA) directs Cas9 (Figure 2) to make a precise double-stranded break in genomic DNA complementary to the guide sequence. The most commonly used CRISPR system is that of *S. pyogenes*, which uses

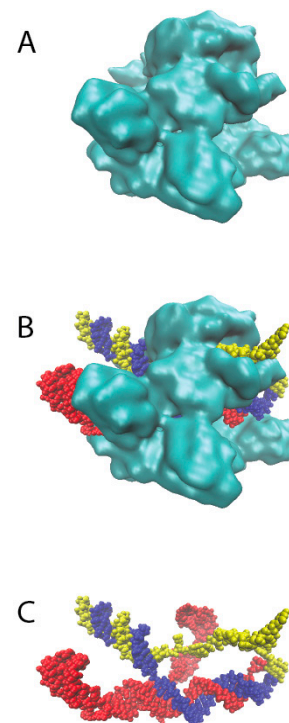


Figure 2. Cas9 structure. A representation of the SpCas9 structure (PDB ID 5Y36) (Huai et al., 2017) in complex with sgRNA (red) and target DNA is shown (B). The complementary and non-complementary strands of the target DNA are indicated in blue and yellow, respectively. The complex with either nucleic acids or SpCas9 removed are shown in panels A and C, respectively.

the *NGG* PAM sequence. This PAM sequence is found roughly every eight base pairs in the human genome, allowing for easy targeting of almost any region of the genome (Cong et al., 2013).

This section of the review details many of the ways researchers have adapted CRISPR-Cas systems to study a wide variety of biological questions. Although CRISPR was first used in human cells only five years ago, CRISPR systems have already been used for genome engineering, pathway component screens, direct regulation of gene expression, imaging of specific DNA sequences, epigenetic modification of specific sequences, and detection of specific nucleotide sequences in body fluid samples. A detailed table containing the CRISPR systems mentioned in this review can be found at the end of this review (Table 1).

The first uses of adapted CRISPR systems were to induce a targeted double-stranded break (DSB) in a DNA sequence (Cong et al., 2013; Jinek et al., 2012). The outcome of this DSB is determined by which repair mechanism is used by the cell to repair the damage (Ran et al., 2013). In eukaryotes, two of the main methods to repair DSBs in DNA are non-homologous end-joining (NHEJ) and homology-directed repair (HDR) (Ran et al., 2013). During NHEJ, the break is reconnected after the removal of overhangs (Ran et al., 2013). There is a small but appreciable chance that

| Application | Use of CRISPR | System Components | Anticipated Result(s) | References |
|--|--|--|--|---|
| Genome engineering | Generation of double-stranded DNA breaks | <ul style="list-style-type: none"> Unmodified spCas9 Gene-specific sgRNA | <ul style="list-style-type: none"> Gene knockouts by NHEJ HDR gene editing | <ul style="list-style-type: none"> Jinek et al., 2012 Cong et al., 2013 Ran et al., 2013 |
| | Generation of single-stranded DNA breaks | <ul style="list-style-type: none"> Cas9 nickase (D10A or H841A) Gene-specific sgRNA | <ul style="list-style-type: none"> HDR gene editing with fewer off-target effects | <ul style="list-style-type: none"> Cong et al., 2013 Mali et al., 2013 |
| | Generation of specific point mutations | <ul style="list-style-type: none"> dCas9 linked to an AID ortholog Gene-specific sgRNA | <ul style="list-style-type: none"> Targeted C to T point mutations | <ul style="list-style-type: none"> Nishida et al., 2016 |
| | | <ul style="list-style-type: none"> dCas9 linked to an adenine base editor Gene-specific sgRNA | <ul style="list-style-type: none"> Targeted A to G point mutations | <ul style="list-style-type: none"> Gaudelli et al., 2017 |
| Screen | Genome-wide screen using double-stranded DNA breaks | <ul style="list-style-type: none"> Unmodified spCas9 sgRNA library | <ul style="list-style-type: none"> Large-scale “drop-out” screens | <ul style="list-style-type: none"> Shalem et al., 2014 Wang et al., 2014 |
| Direct regulation of gene expression | Disrupt gene expression | <ul style="list-style-type: none"> dCas9 (D10A, H841A) Promoter-specific sgRNA | <ul style="list-style-type: none"> Gene knockdown | <ul style="list-style-type: none"> Qi et al., 2013 Chen et al., 2013 |
| | Increase gene expression | <ul style="list-style-type: none"> dCas9 linked to VP64 or SunTag/VP64 Promoter-specific sgRNA | <ul style="list-style-type: none"> Increased gene expression | <ul style="list-style-type: none"> Cheng et al., 2013 Tanenbaum et al., 2014 |
| Imaging | Detection of specific sequences for imaging purposes | <ul style="list-style-type: none"> dCas9 linked to fluorophore | <ul style="list-style-type: none"> Imaging of target DNA sequences | <ul style="list-style-type: none"> Chen et al., 2013 |
| Epigenetic regulation of gene expression | Alter epigenetic modifications | <ul style="list-style-type: none"> dCas9 linked to HAT domain Region-specific sgRNA | <ul style="list-style-type: none"> Increased gene expression | <ul style="list-style-type: none"> Hilton et al., 2015 |
| | | <ul style="list-style-type: none"> dCas9 linked to KRAB domain Region-specific sgRNA | <ul style="list-style-type: none"> Gene silencing | <ul style="list-style-type: none"> Thakore et al., 2016 |
| | | <ul style="list-style-type: none"> dCas9 linked to TET1 domain Region-specific sgRNA | <ul style="list-style-type: none"> Modest increase in gene expression | <ul style="list-style-type: none"> Choudhury et al., 2016 |
| SHERLOCK | Detection of disease-specific DNA in patient samples | <ul style="list-style-type: none"> Cas13a Disease-specific sgRNA | <ul style="list-style-type: none"> Identification of target DNA sequence in sample | <ul style="list-style-type: none"> Gootenberg et al., 2017 |

Table 1. Summary of discussed CRISPR-based systems. CRISPR systems can be adapted for a variety of purposes. Many systems involve the catalytically dead (dCas9), which acts as a targeted DNA binding protein, to which many other effector domains can be added.

NHEJ will result in short insertion or deletion mutations (indels), which can lead to frameshift mutation if they occur in the coding region of a gene (Ran et al., 2013). In HDR, the cell repairs the DSB according to a template with similar flanking sequence, often the same region of the second chromosome in diploid organisms. In CRISPR-based genome engineering, an alternative repair template, with the desired genetic alteration, is introduced (Ran et al., 2013). This template can be either a linearized plasmid for larger alterations or a single-stranded DNA oligonucleotide donor (ssODN) for minor alterations. A ssODN requires only ~40-bp of homology on either side of the cut site for high-efficiency HDR (Ran et al., 2013). Plasmid donors, on the other hand, typically use 500bp of homology on either side of the cut site, but can be used

to insert large segments of DNA, such as fluorescent protein tags (Ran et al., 2013). Both NHEJ and HDR can occur in the same population, particularly if the target gene is nonessential. Transfected cell populations are often clonally diluted and genetically evaluated to prevent unwanted alternate products.

Engineered Cas9 nickase systems can also be used to decrease the probability of NHEJ-induced indel mutations during HDR experiments (Cong et al., 2013). In Cas9 nickase, one of the two nuclease domains are deactivated by a mutation, and the resulting system will only cut one of the two DNA strands. This produces a site-specific single-stranded nick instead of a double-stranded break in the DNA duplex. Work by Feng Zhang and colleagues has shown that the D10A mutant of Cas9 introduced HDR insertions at

the same rate as native Cas9, without producing a single detectable indel mutation (Cong et al., 2013).

In addition to the use of CRISPR systems to study individual genes or mutations, the ease of designing sgRNAs have made CRISPR systems ideal for large-scale screens as well. In early 2014, two papers independently validated the use of massive sgRNA libraries for whole genome screens (Shalem et al., 2014; Wang et al., 2014). One such library, the GeCKO library contains 64,751 sgRNAs which target 18,080 genes in the human genome (Shalem et al., 2014). These libraries can be stably transfected into cells, which are then whole-genome sequenced after selective conditions are applied. The relative abundance of sgRNAs can then be used to determine which genes grant resistance to the experimental condition. “Drop-out” screens have been used to identify genes important for several processes including genes that grant melanoma cells resistance to a therapeutic RAF inhibitor (Shalem et al., 2014), therapeutic targets in acute myeloid leukemia (Tzelepis et al., 2016), oncogenic driver mutations (Kiessling et al., 2016), and genes which affect lung cancer metastasis in mouse models (Chen et al., 2015). Similar screens have been done using RNA interference for years, but CRISPR gives the advantage that expression is interrupted at the DNA level, which enables researchers to interrogate genes that are only natively expressed at low levels. Furthermore, because CRISPR targets the DNA, it can be used to interrogate non-coding regions of DNA such as structural elements and enhancer regions.

The versatility of CRISPR systems extends beyond the ability of Cas9 to create targeted DSBs in DNA. Catalytically deactivated Cas9 (dCas9) can be used as a targeted DNA-binding protein to which functional protein domains can be attached. dCas9 is the result of two point mutations that inactivate the RuvC (D10A) and HNH (H841A) nuclease domains (Qi et al., 2013). The simplest application of dCas9 is CRISPR interference (CRISPRi) (Qi et al., 2013). Although dCas9 does not cleave either DNA strand, it still binds to the DNA and interferes with transcription, leading to lower gene expression levels (Qi et al., 2013). A dCas9 targeted to a transgene for red fluorescent protein (RFP) in *E. coli* resulted in a 2- to 300-fold decrease in observable fluorescence (Qi et al., 2013). The effect was only seen when the sgRNA was targeted to the non-template strand and was dependent on where in the gene the sgRNA was targeted (Qi et al., 2013). Targeting the sgRNA to an upstream region of the gene lead to increased repression (Qi et al., 2013). Furthermore, targeting multiple sgRNAs to the same gene has a multiplicative effect (Qi et al., 2013). Targeting two sgRNAs to the same gene, each of which was found to have a roughly 300-fold repression, resulting in 1,000-fold repression of RFP expression (Qi et al., 2013). This strategy permits tunable repression of gene expression, which would be impossible using catalytically active Cas9. Another study found that they could increase CRISPRi effects by roughly five-fold by redesigning the sgRNA to remove a potential Pol III terminator sequence and elongate the Cas9-binding hairpin structure (Chen et al., 2013).

Deactivated Cas9 can also be used to increase gene expres-

sion by attaching transcriptional activators to dCas9. One study attached a single copy of the activator VP64, a fusion of four copies of the herpes virus transcriptional activator VP16, but found only a modest effect on gene expression (Cheng et al., 2013). A follow-up study found a much larger enhancement using a SunTag protein scaffold to hold ten copies of VP64 on each dCas9 (Tanenbaum et al., 2014). As a proof of concept, the authors targeted the expression of two genes in K562 cells, one not normally expressed (CXCR4) and one normally expressed at high levels (CDKN1B). The SunTag-VP64 increased CXCR4 protein expression by 10-50x, an increase that was sufficient to alter the cell phenotype drastically. The mRNA levels of the cell cycle inhibitor CDKN1B increased by roughly two-fold. This increase was sufficient to significantly reduce cell growth rates (Tanenbaum et al., 2014).

Fluorescent proteins can also be attached to dCas9 to visualize specific DNA sequences. One group used an enhanced GFP (eGFP)-bound dCas9 to study both repeat and non-repeat DNA sequences (Chen et al., 2013). The authors were able to analyze telomere movement during telomere elongation, monitor gene copy number in living cells, and study chromatin architecture during mitosis. They hypothesized that if two different CRISPR systems are used concurrently, CRISPR imaging could be used to visualize the physical interaction of two genetic elements in real time in living cells. The fluorescence signal could be further enhanced by a system like the SunTag with multiple attached fluorescent proteins.

Other groups are studying the use of base-modifier domains linked to dCas9 for single base pair alterations. Akihiko Kondo and colleagues attached an activation-induced cytidine deaminase (AID) ortholog domain onto a CRISPR system which can be used to make targeted point mutations (Nishida et al., 2016). AID creates C to T point mutations by converting the deoxycytidine to a deoxyuridine which is converted to deoxythymidine during DNA replication (Nishida et al., 2016). The authors attached PmCDA, which is an AID ortholog, to dCas9 using a 100 amino acid linker sequence. This complex created point mutations roughly 18 base pairs upstream of the PAM sequence. In *Saccharomyces cerevisiae*, point mutations were observed at the -18bp position 41-51% of the time. When targeted to a position where the adjacent base pairs (-17, -19) were also cytidine, 18-40% mutation rates were observed at these locations as well. However, there was only about a 5% chance of a mutation occurring at the -16 and -20 positions (Nishida et al., 2016). Targeted mutation rates in this system were also improved in yeast by using a Cas9 variant that nicked the DNA strand on the position opposite of the AID activity. The authors suggest that this is because the nick activates a DNA-damage repair mechanism in which the altered deoxyuridine is paired with a deoxythymidine. The Cas9 nickase, however, resulted in unwanted deletion mutations in Chinese Hamster Ovary (CHO) cells. These unwanted deletions could be reduced by the addition of uracil DNA glycosylase inhibitor, which blocks the removal of uracil from DNA, but the rate of indels was still high. These undesired mutations limit the efficacy of the Cas9 nickase-PmCDA1

system for mammalian cells. Another study by David Liu's group used directed evolution of the *E. coli* tRNA adenine deaminase (TadA) to develop a domain which can make targeted A:T to G:C point mutations in DNA when linked to dCas9 (Gaudelli et al., 2017). The resulting protein, which the authors termed Adenine Base Editor (ABE7.10), was able to make targeted A:T to G:C point mutations in HEK293 cells with a 54% efficiency and indel rates below 0.1%. Although these systems are not sufficiently well characterized for clinical applications, the ability to efficiently create targeted point mutations could drastically alter the way genetic diseases are treated.

Other dCas9 systems have been used to make targeted epigenetic changes. Charles Gersbach and colleagues (2015) added a histone acetyltransferase (HAT) domain to a dCas9 and targeted the protein to the promoter region of several different genes (Hilton et al., 2015). The authors found that the dCas9-HAT system was capable of increasing gene expression up to ten-fold as measured by mRNA levels. They also found that just one sgRNA at each location was necessary for activation of the targeted promoter. Gersbach's group (2016) also fused the Kruppel-associated box (KRAB) protein domain to dCas9 for targeted gene silencing (Thakore et al., 2015). They were able to reduce the expression of several globin genes specifically by targeting the dCas9-KRAB system to a distal promoter roughly 10kb away. Joseph Irudayaraj and colleagues fused dCas9 to the catalytic domain of Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) and found it capable of demethylating the BRCA1 gene in HeLa cells (Choudhury et al., 2016). The resulting increase in BRCA1 mRNA levels, however, was modest, a 1.5-2 fold enhancement for the best test samples. Although they are not discussed here, many studies have been published using dCas9 systems for epigenetic modification. For a more detailed review of using CRISPR systems for epigenetic modification, see Brocken et al. (2017). These epigenetic modifications can be used for better interrogation of known epigenetic elements, and to alter cell type to create better disease models.

The final adaptation this review will discuss is the SHERLOCK system, which is a CRISPR-based system for disease detection (Gootenberg et al., 2017). Feng Zhang, James Collins, and colleagues (2017) utilized Cas13a from *Leptotrichia buccalis*, which has targeted endoribonuclease activity. In the SHERLOCK system, DNA samples are transcribed to RNA, which can then be detected by Cas13a (Gootenberg et al., 2017). Interestingly, Cas13a engages in nonspecific cleavage of nearby RNA if the targeted RNA is detected (Gootenberg et al., 2017). SHERLOCK takes advantage of this characteristic by providing labeled RNA which Cas13a nonspecifically cleaves if the targeted DNA is found in the sample. Because this system is CRISPR-based, single base pair mismatches in the PAM region can be used to distinguish between highly similar viral strains, such as Dengue and Zika. This system was also found to be capable of detecting RNA from specific bacteria and could be used to genotype human alleles of known SNPs from saliva samples (Gootenberg et al., 2017). The SHERLOCK system could also detect the presence of EGFR and BRAF

mutations in the cell-free DNA "mock samples", which contained 1 attomolar concentration of ssDNA diluted in a background of genomic DNA (Gootenberg et al., 2017). Lastly, the components of the SHERLOCK system could be produced, freeze-dried, and shipped out for a cost as low as \$0.61 per test, making it a valid tool for disease detection in clinical settings (Gootenberg et al., 2017).

Off-Target Cleavage

One major limitation of Cas9 for both research and medical applications is off-target cleavage, which occurs when Cas9 cleaves a DNA sequence that is not perfectly complementary to its guide RNA. This is problematic because reverse genetics, the process by which specific genes are studied through gene mutation experiments, relies on the assumption that only one gene is mutated in any given experiment. Although significant efforts have been made to characterize Cas9 specificity, off-target cleavage is still not fully understood. These off-target cleavage events could potentially affect any experiment which relies on CRISPR mutations. The risk of off-target cleavage also limits the potential use of CRISPR systems for treatment of disease because off-target cleavage of specific genes could increase the risk of diseases. For example, off-target cleavage of tumor suppressor genes in patients could drastically increase risk of cancer.

Reports generally agree that off-target cleavage is a major concern for Cas9 systems. Cell culture studies have shown that off-target cleavage can be detected at sites with near perfect complementarity with the sgRNA sequences (Cradick et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). It has also been shown that mismatches proximal to the PAM sequence reduce cleavage efficiency more than distal mismatches (Cong et al., 2013; Cradick et al., 2013; Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). In general, it is well accepted that mismatches between the guide RNA and the targeted sequence correlate with reduced cleavage efficiency.

Jeffrey Sander and colleagues, however, have shown that up to five mismatches between the genomic DNA sequence and the sgRNA can be tolerated (Fu et al., 2013). In their study, the authors designed sgRNAs to target specific regions of the VEGFA gene, and then measured off-target cleavage of genomic regions with similar sequences. One of these guide RNAs, targeted to the sequence: GACCCCTCCACCCCGCTCCGG, cleaved the perfectly complementary target site of the VEGFA gene in 50% of the cells tested (Fu et al., 2013). This same guide RNA also cleaved an off-target site in the CALY gene with the sequence: CGCCCTCCCCACCCCGCTCCGG, on a separate chromosome, 44% of the time (Fu et al., 2013). So, although the VEGFA gene was targeted, the CRISPR system cleaved the CALY gene at nearly the same efficiency, despite the four mismatched nucleotides (bolded and underlined) between the two sites. The authors also found that off-target sites can have equivalent, or even greater, rates of cleavage than the on-target site, and that off-target cleavage rates vary significantly between different human cell lines (Fu et al., 2013). Other researchers found a low level of cleavage

in off-target sites that have an *NAG* PAM sequence, suggesting that the *NGG* PAM sequence is not an absolute necessity for Cas9 cleavage (Hsu et al., 2013). Overall, although significant research efforts have been dedicated to fully understanding the nature of off-target Cas9 cleavage, the nature of Cas9 specificity remains poorly understood.

DISCUSSION

Although the first glimpse of CRISPR systems was identified 30 years ago, only within the past 5 years have advanced in our understanding allowed the rapid expansion of CRISPR-based technologies. Interest in CRISPR has grown quickly as its applications have become clear, and there is every indication that the field will continue to expand and develop rapidly. CRISPR and its adaptations have revolutionized genetic engineering, which has drastically altered experimental biology. Significant work remains to characterize native CRISPR biology, improve existing CRISPR systems, and engineer CRISPR systems for alternative uses. Of particular interest are the gaps in understanding of the adaptation phase of native CRISPR systems – specifically understanding how organisms identify invading DNA and incorporate it into their CRISPR arrays. These gaps have not inhibited the development of adapted CRISPR systems but are nonetheless important to understand. For example, scientists studying the use of engineered viruses to treat bacterial infections may find it critical to understand CRISPR immunity systems in order to prevent bacteria from developing immunity to the engineered viruses. Perhaps a Cas1 inhibitor could be used in conjunction with a virus-based treatment to prevent incorporation of viral DNA into the targeted bacterial CRISPR array, a result which would lead to resistance to the treatment.

A great deal of current and future research will focus on improving adapted CRISPR systems. Of particular interest are the characterization of additional Cas nucleases and further study into off-target Cas9 cleavage. Each CRISPR system has its own unique PAM requirements, and Cas9 nucleases from other type II CRISPR systems can be used to access more cut sites in the genome. The *NGG* PAM sequence of *S. pyogenes* Cas9 can be targeted to a cut site roughly every eight base pairs. This is insufficiently precise for some gene editing purposes, such as AID-based targeted point mutations. The lab of Virginijus Siksnys recently developed a new library-based technology which can be used to determine the PAM specificity of newly discovered Cas9 proteins from different organisms (Karvelis et al., 2015). The authors then used this new technique to characterize the PAM sequence of the *Brevibacillus laterosporus* bacterium, which was found to be NNNNCNDD, in which N corresponds to any nucleotide and D corresponds to A, G, or T (Karvelis et al., 2015). Usage of this newly characterized Cas9 and discovery of additional Cas9 proteins from other type II CRISPR systems will increase the portion of the genome accessible to CRISPR targeting. Other researchers are also working on mutating the spCas9 protein to alter its PAM specificity. The group of Keith Joung was able to use both rational mutation and directed evolution in a bacterial system to create spCas9 variants,

which specifically cleave target sites with *NGA*, rather than *NGG*, PAM sequences (Kleinstiver et al., 2015). Additional directed evolution experiments, performed with spCas9 as well as other Cas9 variants, could be used to design a series of Cas9 variants, each with different PAM specificities, and therefore capable of targeting unique sites inaccessible to other Cas9 variants.

Additionally, off-target effects have been noted as a serious concern, particularly if CRISPR is to be developed into a therapeutic disease treatment. CRISPR-based HDR has enormous potential for the treatment of genetic diseases, and even low rates of off-target Cas9 cleavage could prevent clinical applications of this technology. To solve this problem, researchers are again turning toward engineered Cas9 variants. In 2016, two research groups used rational mutation to design spCas9 variants which have reduced off-target cleavage activity (Kleinstiver et al., 2016; Slaymaker et al., 2016). Although each research group chose different amino acids in spCas9 to mutate, they each achieved lower off-target cleavage rates while maintaining high on-target cleavage activity (Kleinstiver et al., 2016; Slaymaker et al., 2016). Additional work will be required to confirm these results and fully characterize the activity of these Cas9 variants, both in human cells and *in vivo*.

Furthermore, other effector domains can be attached to dCas9 for additional DNA-targeted functions. Additional epigenetic writer and eraser domains are promising candidates for these effector domains. There are dozens of post-translational modifications (PTMs) which can be added to histone tails to regulate gene expression and function. The function of many of these epigenetic marks remains unclear. So far, CRISPR-Cas9 systems have been developed that alter the histone tail acetylation and methylation state. The dCas9-HAT system increases gene expression by adding acetyl marks to histone tails, and the dCas9-KRAB system represses gene expression in part by inducing the methylation of histone tails in gene enhancer regions (Hilton et al., 2015; Thakore et al., 2015). However, histone tails can also be phosphorylated, ADP-ribosylated, ubiquitylated, and sumoylated on numerous different amino acid positions (Rothbart & Strahl, 2014). The function of many of these histone marks, as well as how different marks interact with each other, are poorly understood. Attaching writer and eraser domains for each of these epigenetic marks to dCas9 can be used for the targeted addition or removal of these uncharacterized marks, which will likely prove crucial in the future study of histone PTMs.

CONCLUSION

CRISPR-Cas systems are adaptive prokaryotic immune systems which have been repurposed for genome editing purposes. In the native organism, these systems incorporate portions of invading DNA into the host organism's genome, granting the organism resistance to subsequent infections. The study of these CRISPR systems has led to the discovery of remarkably useful proteins, such as the Cas9 nuclease of *S. pyogenes*. Adapted CRISPR systems based on the Cas9 nuclease can be used to create targeted double-stranded breaks in nearly any location of the genome of every

organism studied so far. Additional applications can be achieved using dCas9 systems, which convert Cas9 into a targeted DNA-binding protein. These dCas9 systems have been used to visualize DNA sequences, alter gene expression, and make targeted epigenetic alterations. The versatility and ease of CRISPR systems have made them revolutionary, and future improvements will continue to add to their value.

ACKNOWLEDGMENTS

The authors thank Towson University for support during this research effort.

REFERENCES

- Addgene. (2017). CRISPR 101: A desktop resource. Retrieved from <https://info.addgene.org/download-addgenes-ebook-crispr-101-2nd-edition>
- Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S., & Shinkai, A. (2010). Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *Journal of Molecular Biology*, 395(2), 270–281. <https://doi.org/10.1016/j.jmb.2009.10.057>
- Arslan, Z., Wurm, R., Brenner, O., Ellinger, P., Nagel-steger, L., Oesterhelt, F., ... Willbold, D. (2013). Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Research*, 41(12), 6347–6359. <https://doi.org/10.1093/nar/gkt315>
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709–1712.
- Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extra-chromosomal origin. *Microbiology*, 151, 2551–2561. <https://doi.org/10.1099/mic.0.28048-0>
- Brocken, D. J. W., Tark-Dame, M., & Dame, R. T. (2017). dCas9: A Versatile Tool for Epigenome Editing. *Current Issues in Molecular Biology*, 26(Sep 7), 15–32. <https://doi.org/10.21775/cimb.026.015>
- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G., ... Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7), 1479–1491. <https://doi.org/10.1016/j.cell.2013.12.001>
- Chen, S., Sanjana, N. E., Zheng, K., Shalem, O., Lee, K., Shi, X., ... Sharp, P. A. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, 160(6), 1246–1260. <https://doi.org/10.1016/j.cell.2015.02.038>
- Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., ... Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Research*, 23(10), 1163–1171. <https://doi.org/10.1038/cr.2013.122>
- Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B., & Irudayaraj, J. (2016). CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget*, 7(29), 46545–46556. <https://doi.org/10.18632/oncotarget.10234>
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–824. <https://doi.org/10.1126/science.1231143>
- Cradick, T. J., Fine, E. J., Antico, C. J., & Bao, G. (2013). CRISPR/Cas9 systems targeting b-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Research*, 41(20), 9584–9592. <https://doi.org/10.1093/nar/gkt714>
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., ... Eckert, M. R. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602–607. <https://doi.org/10.1038/nature09886>
- Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), 1077–1087. <https://doi.org/10.1126/science.1258096>
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, K., & Sande, J. D. (2013). High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 9(31), 822–826. <https://doi.org/10.1038/nbt.2623>
- Fusco, S., Liguori, R., Limauro, D., Bartolucci, S., She, Q., & Contursi, P. (2015). Biochimie transcriptome analysis of *Sulfolobus solfataricus* infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system. *Biochimie*, 118, 322–332. <https://doi.org/10.1016/j.biochi.2015.04.006>
- Garneau, J. E., Fremaux, C., Horvath, P., & Magada, A. H. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), 67–72. <https://doi.org/10.1038/nature09523>
- Gasiunas, G., Barrangou, R., Horvath, P., Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Immunity in bacteria Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS*, 109(39), 15539–15548. <https://doi.org/10.1073/pnas.1208507109>
- Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, 551(7681), 464–471. <https://doi.org/10.1038/nature24644>
- Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., ... Koonin, E. V. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, 442, 438–442. <https://doi.org/10.1126/science.aam9321>
- Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Computational Biology*, 1(6), 474–483. <https://doi.org/10.1371/journal.pcbi.0010060>
- Heler, R., Samai, P., Modell, J. W., Weiner, C., Gregory, W., Bikard, D., & Marrafini, L. A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature*, 519(7542), 199–202. <https://doi.org/10.1038/nature14245>
- Hilton, I. B., Ippolito, A. M. D., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, 33(5), 510–519. <https://doi.org/10.1038/nbt.3199>
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 9(31), 827–832. <https://doi.org/10.1038/nbt.2647>
- Huai, C., Li, G., Yao, R., Zhang, Y., Cao, M., Kong, L., ... Huang, Q. (2017). Structural insights into DNA cleavage activation of CRISPR-Cas9 systems. *Nature Communications*, 8, 1375–1384. <https://doi.org/10.1038/s41467-017-01496-2>
- 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. *Journal of Bacteriology*, 169(12), 5429–5433.
- Jansen, R., Embden, J. D. A. Van, Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565–1575.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816–822. <https://doi.org/10.1126/science.1225829>
- Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., ... Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343(6176), 1215–1226. <https://doi.org/10.1126/science.1247997>
- Ka, D., Lee, H., Jung, Y., Kim, K., Seok, C., Suh, N., & Bae, E. (2016). Crystal Structure of *Streptococcus pyogenes* Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System Article Crystal Structure of *Streptococcus pyogenes* Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure/Folding and Design*, 24(1), 70–79. <https://doi.org/10.1016/j.str.2015.10.019>
- Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., & Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biology*, 16(253). <https://doi.org/10.1186/s13059-015-0753-1>

- org/10.1186/s13059-015-0818-7
- Kiessling, M. K., Schuierer, S., Stertz, S., Beibel, M., Bergling, S., Knehr, J., ... Roma, G. (2016). Identification of oncogenic driver mutations by genome-wide CRISPR-Cas9 dropout screening. *BMC Genomics*, 17(1), 1–16. <https://doi.org/10.1186/s12864-016-3042-2>
- Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), 490–495. <https://doi.org/10.1038/nature16526>
- Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., ... Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523(7561), 481–485. <https://doi.org/10.1038/nature14592>
- Lander, E. S. (2015). The heroes of CRISPR. *Cell*, 164(1–2), 18–28. <https://doi.org/10.1016/j.cell.2015.12.041>
- Makarov, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Mojica, F. J. M., Wolf, Y. I., ... Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, 9, 467–478.
- Makarov, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., ... Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, 13, 722–736. <https://doi.org/10.1038/nrmicro3569>
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., ... Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, 339(6121), 823–827. <https://doi.org/10.1126/science.1232033>
- Mojica, F. J. M., Diez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3), 733–740. <https://doi.org/10.1099/mic.0.023960-0>
- Mojica, F. J. M., Diez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Molecular Evolution*, 60(2), 174–182. <https://doi.org/10.1007/s00239-004-0046-3>
- Morange, M. (2015). What history tells us XXXVII. CRISPR–Cas: The discovery of an immune system in prokaryotes. *Journal of Biosciences*, 40(2), 221–223. <https://doi.org/10.1007/s12038-015-9532-6>
- Nam, K. H., Ding, F., Haitjema, C., Huang, Q., Delisa, M. P., & Ke, A. (2012). Double-stranded Endonuclease Activity in *Bacillus halodurans* Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) -associated Cas2 Protein. *Journal of Biological Chemistry*, 287(43), 35943–35952. <https://doi.org/10.1074/jbc.M112.382598>
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., ... Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, 353(6305), 1248–1256. <https://doi.org/10.1126/science.aaf8729>
- Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., ... Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, 156(5), 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>
- Núñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, 527(7579), 535–538. <https://doi.org/10.1038/nature15760>
- Núñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1 – Cas2 complex formation mediates spacer acquisition during CRISPR – Cas adaptive immunity. *Nature Structural & Molecular Biology*, 21(6), 528–536. <https://doi.org/10.1038/nsmb.2820>
- 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. *Nature Biotechnology*, 9(31), 839–843. <https://doi.org/10.1038/nbt.2673>
- Patterson, A. G., Yevstigneyeva, M. S., & Fineran, P. C. (2017). Regulation of CRISPR – Cas adaptive immune systems. *Current Opinion in Microbiology*, 37(0), 1–7. <https://doi.org/10.1016/j.mib.2017.02.004>
- 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>
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), 1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR–Cas9 system. *Nature Protocols*, 8(11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>
- Rothbart, S. B., & Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta*, 1839(8), 627–643. <https://doi.org/10.1016/j.bbagr.2014.03.001>
- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR / Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research*, 39(21), 9275–9282. <https://doi.org/10.1093/nar/gkr606>
- Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T., ... Zhang, F. (2014). Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science*, 343(6166), 84–87. <https://doi.org/10.1126/science.1247005>
- Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351(6268), 84–89. <https://doi.org/10.1126/science.aad5227>
- Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., & Vale, R. D. (2014). A Protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, 159(3), 635–646. <https://doi.org/10.1016/j.cell.2014.09.039>
- Thakore, P. I., Song, L., Safi, A., Shivakumar, K., Kabadi, A. M., Reddy, T. E., ... Gersbach, C. A. (2015). Highly specific epigenome editing by CRISPR/Cas9 repressors for silencing of distal regulatory elements. *Nature Methods*, 12(12), 1143–1149. <https://doi.org/10.1038/nmeth.3630>
- Tzelepis, K., Koike-yusa, H., Braekeleer, E. De, Pina, C., Vassiliou, G. S., Li, Y., ... Yusa, K. (2016). A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in Acute Myeloid Leukemia. *Cell Reports*, 17(4), 1193–1205. <https://doi.org/10.1016/j.celrep.2016.09.079>
- van der Oost, J., Westra, E. R., Jackson, R. N., & Wiedenheft, B. (2014). Unraveling the structural and mechanistic basis of CRISPR–Cas systems. *Nature Reviews Microbiology*, 7(12), 479–492. <https://doi.org/10.1038/nrm.2014.371>
- Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic screens in human cells using the CRISPR–Cas9 system. *Science*, 343(6166), 80–85. <https://doi.org/10.1126/science.1246981>
- Wei, Y., Terns, R. M., & Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR – Cas adaptation. *Genes & Development*, 29(4), 356–361. <https://doi.org/10.1101/gad.257550.114>
- Wright, A. V., Liu, J.-J., Knott, G. J., Doxzen, K. W., Nogales, E., & Doudna, J. A. (2017). Structure of the CRISPR genome integration complex. *Science*, 357(6356), 1113–1118. <https://doi.org/10.1126/science.aao0679>
- Young, J. C., Dill, B. D., Pan, C., Hettich, R. L., Banfield, J. F., Shah, M., ... Verberkmoes, N. C. (2012). Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLOS ONE*, 7(5). <https://doi.org/10.1371/journal.pone.0038077>