ABSTRACT

"CRISPR: Its Origins, Functions and Future"

Priyankaa Anna Bhatia

Director: Michael A. Trakselis, Ph.D.

What started as simply an observation of a confusing set of tandem palindromic repeats in bacterial genomes grew into the discovery and harnessing of the CRISPR-Cas9 mechanism of gene editing. Clustered regularly interspaced short palindromic repeats (CRISPR) are a naturally occurring adaptive immune system present in microbes. They direct foreign DNA cleavage by storing DNA sequences from the foreign invaders and then using those sequences in a protein effector complex to bring about double stranded breaks. They are mainly used against bacteriophages. Once this mechanism was discovered and studied in its various forms, it was found that the Cas9 complex could be used in other organisms and could be engineered to target specific DNA sequences. This opened the floodgates for research into the gene editing properties of CRISPR-Cas9 and its application to therapeutic genetics as well as drug research. The perfection of the technology brought along with it a long running patent battle between two labs for the rights to the technology as well as a constant emergence of new research showing the benefits, possibilities and problems CRISPR-based editing brings to light.

APPROVED BY DIRECTOR OF HONORS THESIS

_____

Dr. Michael A. Trakselis, Department of Chemistry

APPROVED BY THE HONORS PROGRAM

_____

Dr. Elizabeth Corey, Director

DATE: _____

CRISPR: ITS ORIGINS, FUNCTIONS AND FUTURE

A Thesis Submitted to the Faculty of

Baylor University

In Partial Fulfilment of the Requirements for the

Honors Program

By

Priyankaa Bhatia

Waco, Texas

May 2017

TABLE OF CONTENTS

Table of Figures iii

Chapter One: CRISPR-How It All Began 1

Chapter Two: Classifying CRISPR systems 15

Chapter Three: The Mechanism of Cas9 and Cascade Systems 29

Chapter Four: Harnessing the Tool 41

LIST OF FIGURES

Figure 1.1: Overview of immunity 6

Figure 2.1: 2011 Classification 16

Figure 2.2: 2015 Classification 17

Figure 2.3: Cas3 structure 18

Figure 2.4: Type I subtypes 19

Figure 2.5: Ribbon structure of Cas10 20

Figure 2.6: Type II subtypes 20

Figure 2.7: Type IV subtype 21

Figure 2.8: Type II subtypes 22

Figure 2.9: Archaea Cas9 23

Figure 2.10: Type V system and *cpf1* domain layout 24

Figure 2.11: Type V subtypes 25

Figure 3.1: Ribbon structure of catalytic core of Cas9 30

Figure 3.2: Ribbon structure of SpyCas9 31

Figure 3.3: *cas9* domains 31

Figure 3.4: Mechanism of Cas9 action 32

Figure 3.5: Model of protospacer incorporation 34

Figure 3.6: Model of crRNA maturation 35

Figure 3.7: Gene sequence of Cascade complex 36

Figure 3.8: Model of Cascade bound to dsDNA 37

Figure 3.9: Target recognition and binding by Cascade 38

Figure 4.1: Types of double stranded repair 44

CHAPTER ONE

CRISPR: How it All Began

Discovery of Palindromic Repeats

In December, 1987 a paper published in the Journal of Bacteriology investigated the *iap* gene in *Escherichia coli* and set out to characterize its function as an isozyme associated with the membrane.[1] The researchers documented five unusual homologous sequences arranged as repeats flanking the 3' end of the *iap* gene. Thinking nothing much of it, they closed the paper with a line stating they had not found any homologues to this sequence elsewhere and had no clue to its biological significance. Little did these researchers know that they had stumbled upon what was later dubbed CRISPR–Clustered Regularly Interspaced Short Palindromic Repeats–an innate immunity system present in bacteria and archaea that provided resistance to invading phages. It would take another eight years for such a repeat to be noticed again, and another three years for its significance to be fully realized.

In August of 1993, a paper emerged from the University of Alicante in Spain in which Francisco Mojica, a microbiologist who was studying *Haloferax mediterranei* in Santa Pola's marshes, found palindromic repeating sequences in the genome of the archaea.[2] In his 1993 paper, he simply sequenced the DNA and identified the repeating structures. However, his curiosity was piqued and he continued to investigate this curious phenomenon. In July 1995, he published a paper where he sequenced the whole stretch of these tandem repeats and tried to investigate a possible biological role for them.

Mojica hypothesized that these palindromic repeats had some role to play in transcription or replication regulation. After conducting knockout experiments with the repeats, he was only able to conclude that his results "strongly favored" the idea that the tandem repeats played a role in the cell's life cycle.[3] From 1995 to around 2002, many more reports of these repetitive sequences began to emerge and Mojica himself found them in around 20 different microbes.[2] A name was finally given to these sequences in 2002 by Jansen *et al*.[4] For convenience and consistency, they stated that they would refer to the repeats as CRISPR. The name stuck and discoveries of CRISPR loci continued to grow.

But What Does It Do?

In spite of more CRISPR loci being found in different organisms and species, researchers were still not able to pin down their biological function. Such a widespread, conserved genetic phenomenon must be of great biological significance for it to pervade in both bacteria and archaea. Mojica had a breakthrough after running the spacer sequences in genome databases and found that some of the spacers matched known sequences from certain bacteriophages. After running more spacers through the databases, he was able to speculate about the role of CRISPR in providing immunity for the bacteria.[5] Similarly, Vergnaud[6] and Bolotin[7] also published papers describing the extrachromosomal origin of the spacers and attributing them to some kind of immunity for the organism. However, it was only in 2005 that Phillipe Horvath and his team working at Danisco were able to show experimental evidence for the theory that CRISPR gives cells adaptive immunity.

Organisms have many different types of immune systems to protect themselves from pathogens in their environment. Some systems consist of large, physical barriers, like skin, while others are at the molecular level and are mobilized once the organism has been invaded. Innate immunity is a genetically programed set of responses that are made active immediately upon infection. However, innate immunity does not store any memory of the pathogens. Adaptive immunity, on the other hand, adapts to a specific pathogen and changes with the nuances of the ongoing infection. CRISPR appeared to be a system of innate adaptive immunity.[8]

While trying to find a way to prevent dairy bacteria from attack by viruses, the group at Danisco ended up investigating the CRISPR system.[9] They understood that the spacers in between the palindromic repeats were derived from the bacteriophages, and that the *cas* (CRISPR-associated) genes played a role in this resistance. They attacked various strains of bacteria with bacteriophages and observed how the spacers of each differed depending on the phage used. This way, they were able to conclude with certainty that the CRISPR system is adaptive and changes rapidly with exposure to phages. Additionally, they were able to create CRISPR mutants and knock out the *cas7* and *cas9 (*referred to as *cas5* in the paper) genes in order to derive some understanding of their roles in immunity. They found that without *cas7* the bacteria were not able to generate phage resistant mutants and concluded that *cas7* (and *cas9)* were clearly important to the entire process of immunity. It was still unknown what role these *cas* genes played. While many questions still lingered, information about this fascinating immunity system was growing and researchers were beginning to see that this was just the tip of the iceberg.

Cas9 and the PAM sequence

Bolotin *et al*., in their 2005 paper, identified and characterized some of the *cas* genes, including the important *cas9* gene (then identified as *cas5*).[7] They were able to identify "Cas5" as the protein that degraded the DNA of the phage invader. They also identified a commonality between the spacer sequences that all ended with a short 5bp sequence and the corresponding viral genes. It was only in February of 2008 that Deveau *et al*. were able to look into these common sequences and recognize them as important to identifying the spacer on the phage's genome for conferring immunity.[10] Similarly, in the same month, Horvath *et al*. identified these sequences and stated their importance in guiding the CRISPR system.[11] From this 5 bp sequence, they discovered a conserved repeating three base motif, which became known as the proto-spacer adjacent motif, or PAM site.

Going from Gene to Immunity

Up until 2007, researchers had a general understanding of the role of CRISPR and some its basic components. However, they were still not clear on the details of its mechanism and the role of the *cas* genes associated with the repeats. It constituted immunity for the cell, but how exactly go about neutralizing the invading phage? An August 2008 paper by Koonin and Van der Oost was significant in understanding targeting and the complete process of CRISPR.[12] They studied the *Cascade* (CRISPR-associated complex for antiviral defense) complex of *E. coli*. "Cascade" was the name given for the eight *cas* genes and associated spacers that constitute the antiviral defense system of *E. coli*. Their examination of the Cascade complex led them to discover that the immune response is mediated by a crRNA (CRISPR-RNAs), a section of RNA

transcribed out of the spacer region in the CRISPR gene, and that this crRNA was only transcribed when the entire Cascade complex was present. They concluded from their findings and artificial CRISPR design that crRNA targeted the phage and was the link between the gene and immune response. Additionally, they created the artificial CRISPR in both the sense and anti-sense directions. Even though both worked at varying levels, the sense strand yielded the highest results, suggesting mRNA couldn't be the target, which led to the important hypothesis that the crRNA targets the DNA of the phage and exerts a response by interfering with the DNA, not the RNA (as was previously thought). However, this was still a hypothesis and had not yet been tested directly.

Marraffini and Sontheimer in December of 2008 published a paper definitively concluding and experimentally showing that crRNA does in fact target the DNA of the invading phage. They modified one spacer in *S. epidermidis* and included an intron in it which, if it resulted in no cleavage, would suggest DNA was the target of the transcribed crRNA. This finding allowed them to conclude that CRISPR functions very differently from the RNA Interference (RNAi) mechanism (which until this point, researchers considered to be an analogous mechanism). Interestingly, Marraffini and Sontheimer were also the first researchers to explicitly acknowledge and directly state the programmable capabilities of CRISPR. They likened it to a programmable restriction enzyme and stated that CRISPR "can be programmed by a suitable effector crRNA. If CRISPR interference could be manipulated in a clinical setting, it would provide a means to impede the ever-worsening spread of antibiotic resistance genes and virulence factors in staphylococci and other bacterial pathogens."[13] Figure 1.1 represents an overview of the process of immunity that CRISPR brings about.

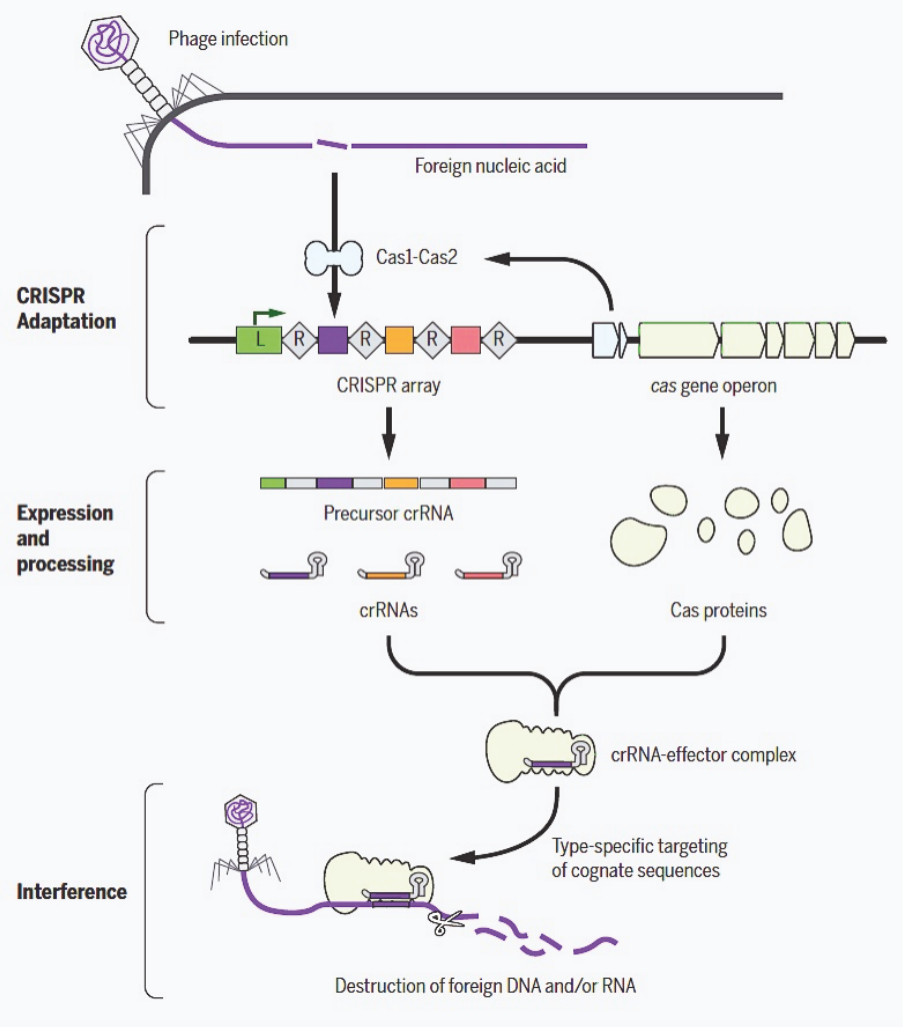


Figure 1.1: Illustration of the three main stages of CRISPR/Cas defense. Adaption is when the phage initially invades and the bacteria includes the foreign DNA into its genome. Expression occurs when the virus invades again and the crRNAs are transcribed from the spacer sequences. Interference occurs with the Cas effector complexes and destroys the viral DNA or RNA.[14]

Cas9 and Double Stranded Breaks in DNA

Scientists had now gathered enough information to understand the basis of what CRISPR does in bacteria. Unfortunately, CRISPR worked so quickly *in vivo* that it was difficult to isolate the mechanism of attack and observe how the target DNA was being destroyed. However, in 2010, Moineau and his colleagues caught a break when they discovered a strain of *S. thermophilus* that was only bringing about partial breakdown of unwanted plasmids (some of the CRISPR systems they were studying not only attacked invading phages, but also targeted certain plasmids.)[2] Thus, they were able to examine the damaged DNA of the plasmids and found that the CRISPR system was causing a double strand break. They turned to the DNA of the bacteriophages, and using PCR and sequencing techniques, were able to confirm that CRISPR also caused double stranded breaks here. They found that CRISPR was causing a blunt end cleavage "three bases upstream of the proto-spacer adjacent motif, as observed for the linearized plasmid" and were able to confirm a similar pattern of cleavage in the phage DNA as well. They also confirmed that the nuclease activity depended solely on the Cas5 [Cas9] protein and that the cleavage that was occurring was "proto-spacer specific and orientation dependent."[15]

Finding the tracrRNA

In 2010, when Emmanuelle Charpentier and Jörg Vogel were classifying microbial RNA, they were not even remotely investigating CRISPR in bacteria. However, their research led them to find a highly prevalent and common class of RNA being transcribed in bacteria—a strain coming from a sequence right next to the CRISPR loci.[2] This sequence was not a *cas* sequence and was not translated into protein. Their

curiosity was piqued and they began to investigate what role this RNA plays in CRISPR, if it plays a role at all. Coined tracrRNA (*trans*-activating CRISPR RNA), the sequence yielded RNAs that had a 25-nucleotide sequence complementary to all repeats of CRISPR in the cell. Their experiments led them to conclude that the tracrRNA was one of the first examples to be found of non *cas*-proteins being recruited for the function of CRISPR. It appeared to function in tandem with the transcribed crRNA and bring about the maturation of crRNA from pre-crRNA.[16] The tracrRNA was responsible for refining the crRNA to match the protospacer unit on the target DNA.

Later on, in 2012, Charpentier and her team worked on refining their previous hypothesis and purifying out the Cas9 protein alone to see what components are required for its function.[17] They found that there was enhanced DNA targeting when the tracrRNA was included to the Cas9 complex along with the crRNA and concluded that the tracrRNA was not only needed for the maturation of the crRNA, but also for the proper orientation of the complex on the target DNA. Further experimentation allowed them to end their paper with some very important findings to the mechanism of Cas9 and the programmable potential of CRISPR. They stated that the Cas9 protein functions as an endonuclease to cleave the target DNA and is guided by both crRNA and tracrRNA. The crRNA contains the sequence that matches a portion of the DNA downstream from a PAM site, meaning crRNA is part of the alterable mechanism for targeting any gene on DNA. They state that their study leads to the "exciting possibility of developing a simple and versatile RNA-directed system to generate dsDNA breaks for genome targeting and editing."[17]

CRISPR Transplanted and Studied In Vitro

The idea that CRISPR could be engineered and used for genome editing in organisms was exciting. However, for this to become a reality, researchers needed to first find out if it was even possible for CRISPR to function outside of its native context. In 2011, Siksnys *et al*. wanted to investigate whether it would be possible to transplant a particular CRISPR system from one organism into another and test if it confers immunity.[18] To do this, they used the Type II CRISPR system in *S. thermophilus* and tested if it could work in *E. coli,* which uses a Type I system. They transformed the *E. coli* genome to include the *S. thermophilus* CRISPR locus and cloned the transformed *E. coli*. The CRISPR system they inserted managed to provide interference against plasmids and phages whose sequences they had inserted as spacers. Excitingly, this would allow for beginning the development of a modifiable CRISPR engineering system. Additionally, Siksnys was able to prove that Cas9 is the only protein required to provide full resistance in another organism if the spacer sequence is already included. Since Cas9 contains a HNH domain, which is common to nucleases, they concluded that the HNH domain cleaves the DNA and is the only thing required to bring about the double stranded breaks, suggesting that immunity could be conferred by only splicing in the *cas9* gene, the leader sequence, and the spacer sequence. Now that CRISPR had been shown to work in another organism, could it be isolated and studied in test tubes?

Following this paper, Siksnys and his group added on to their research by using the same *E. coli* expressing the *S. thermophilus* CRISPR/Cas system to isolate and purify the Cas9-crRNA complex.[19] They then took the purified complex and combined it *in vitro* with oligonucleotides matching the protospacer unit. The purified system brought

about immunity and cleaved all the oligonucleotides that contained the PAM sequence. They then went one step further and engineered a unique crRNA sequence into the CRISPR locus and created a plasmid containing the complimentary sequence. They purified this modified complex and studied it *in vitro* as well. The plasmids containing the complimentary sequence were cleaved by the CRISPR complex, showing that it is in fact, a modifiable genetic technology that can be harnessed by researchers. Around the same time, Charpentier teamed up with Jennifer Doudna to also explore the programmable aspects of a CRISPR-mediated immune response.[17] They not only showed the function of CRISPR *in vitro* and explored the ability to program the tracrRNA, but went one step further to see if it was possible to create a single RNA-guided Cas9 complex that could be constructed to have the function of both the crRNA and the tracrRNA. They tested their hypothesis by engineering five different chimeric RNA guides to work against different targets and found that in all the cases, the Cas9 complex effectively cleaved the DNA at the correct target site. This find allowed them to conclude that "The system is efficient, versatile, and programmable" and "could lead to exciting prospects of genetic engineering in the future."[17]

Editing in Mammalian Cells

From the time that CRISPR was found to be a gene editing mechanism, it was clear that efforts would be made to find its potential for altering mammalian genomes. Up until this point, researchers conducted genome related research on eukaryotic cells, plants or mice, by altering various aspects of their genes using techniques such as zinc-finger nucleases (ZFNs) and transcription activating proteins called transcription

activator-like effector nucleases (TALENs). While these were effective, they were tedious and not foolproof.[2] The harnessing of a tool like CRISPR would change the way gene research was conducted and would pave a way to working on cancer and genetic disorders. However, the largest question first had to be answered: would Cas9 editing complexes work in a mammalian cell?

Feng Zhang, working at the Broad Institute at MIT, set out in 2011 to see if it would be possible to get Cas9 editing to work on mammalian cells. He had minimal effects early on and went about optimizing the system.[2] In February of 2013, Zhang and his research group published a paper showing that not only did a Type II CRISPR/Cas9 system bring about double stranded breaks to DNA in mammalian cells, but that it was possible to enable precise homology-directed repair and encode multiple guide sequences into a single CRISPR array, thus facilitating multiple edits from one complex.[20] Zhang harnessed the *S. pyogenes* type II CRISPR system and designed a spacer to target a locus in human kidney cells that was the appropriate distance from a PAM site. They found that while the four component CRISPR system (RNase III, tracrRNA, pre-crRNA and Cas9) did yield the cleavage, a three-component system without RNase III also worked equally as efficiently. They were also able to mimic the chimeric RNA (chiRNA) that Charpentier and Doudna explored and found that while not as efficient as the RNA duplex, the chiRNA did yield cleavage in certain genomic targets. Then they modified the Cas9 gene to create a protein that nicked, not cleaved, the DNA—allowing for high-fidelity homology-directed repair to take place. In one paper, Zhang and his colleagues were able to address many of the questions and hypotheses that researchers had about the ability for CRISPR to work in mammalian cells.

In the same issue of *Science,* following Zhang's paper, George Church and his research team also published a paper that explored Cas9 genome editing in human cells.[21] Their study was similar to Zhang's in that it also proved that Cas9 could be used in human cells. However, Church and his group engineered a "human codon-optimized version" of the Cas9 protein and created a longer single guide RNA to lead the cleavage. They found better cleavage from the longer single RNA and were able to bring about gene correction by introducing a repair donor strand of DNA to act a template for the repair.

Following these discoveries of the mammalian editing capabilities of CRISPR, research into the potentials and manipulability of CRISPR increased drastically, rapidly broadening the scope of the field. Zhang submitted and was granted a patent for the Cas9 technology in 2013. However, Charpentier and Doudna had filed a patent prior to Zhang's, which was awaiting approval at the time, regarding the ability to use Cas9 in other prokaryotic organisms. The subsequent patents and battle over them will be discussed in a later chapter. The research into the function and mechanism of CRISPR was, in more recent years, also characterized by the discovery of many alternate forms of the system as well as structural studies into the characteristics of the Cas9 gene.

[1] Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. Nucleotide Sequence of the *Iap* Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in *Escherichia Coli*, and Identification of the Gene Product. *J Bacteriol*. **1987**, *169* (12), 5429–5433.

[2] Lander, E. S. The Heroes of CRISPR. *Cell* **2016**, *164* (1–2), 18–28.

[3] Mojica, F. J. M.; Ferrer, C.; Juez, G.; Rodríguez-Valera, F. 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*. **1995**, *17* (1), 85–93.

[4] Jansen, R.; van Embden, D. A. Jan; Gaastra, W.; Schouls, L. M. Identification of Genes That Are Associated with DNA Repeats in Prokaryotes. *Mol. Microbiol*. **2002**, *43* (6), 1565–1575.

[5]Mojica, F. J. M.; Díez-Villaseñor, C.; Soria, E.; Juez, G. Biological Significance of a Family of Regularly Spaced Repeats in the Genomes of Archaea, Bacteria and Mitochondria. *Mol. Microbiol*. **2000**, *36* (1), 244–246.

[6] Pourcel, C.; Salvignol, G.; and Vergnaud, G. CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies. *Microbiology* **2005,** 151 (3), 653–663.

[7] Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D. Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin. *Microbiology* **2005**, *151* (8), 2551–2561.

[8] Parham, P. *The Immune System*, 4 Edition.; Garland Science: New York, NY, 2014.

[9]Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P. CRISPR Provides Acquired Resistance against Viruses in Prokaryotes. *Science* **2007**, *315* (5819), 1709–1712.

[10] Deveau, H.; Barrangou, R.; Garneau, J. E.; Labonté, J.; Fremaux, C.; Boyaval, P.; Romero, D. A.; Horvath, P.; Moineau, S. Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J Bacteriol* **2008**, *190* (4), 1390–1400.

[11] Horvath, P.; Romero, D. A.; Coûté-Monvoisin, A.-C.; Richards, M.; Deveau, H.; Moineau, S.; Boyaval, P.; Fremaux, C.; Barrangou, R. Diversity, Activity, and Evolution of CRISPR Loci in *Streptococcus thermophilus*. *J. Bacteriol*. **2008**, *190* (4), 1401–1412.

[12] 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, D. O. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **2008**, *321* (5891), 960–964.

[13] Marraffini, L. A.; Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA. *Science* **2008**, *322* (5909), 1843–1845.

[14] Jackson, S. A.; McKenzie, R. E.; Fagerlund, R. D.; Kieper, S. N.; Fineran, P. C.; Brouns, S. J. J. CRISPR-Cas: Adapting to Change. *Science* **2017**, *356* (6333), eaal5056.

[15] Garneau, J. E.; Dupuis, M.-È.; Villion, M.; Romero, D. A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadán, A. H.; Moineau, S. The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA. *Nature* **2010**, *468* (7320), 67–71.

[16] Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III. *Nature* **2011**, *471* (7340), 602–607.

[17] Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821.

[18] Sapranauskas, R.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. The *Streptococcus thermophilus* CRISPR/Cas System Provides Immunity in *Escherichia coli*. *Nucleic Acids Res*. **2011**, *39* (21), 9275–9282.

[19] Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas9–crRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria. *PNAS* **2012**, *109* (39), E2579–E2586.

[20] Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; *et al*. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339* (6121), 819–823.

[21] Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339* (6121), 823–826.

CHAPTER TWO

Classifying CRISPR Systems

Mode of Classification

Classifying the various CRISPR-cas loci is a very difficult task because there is a large amount of variability in the genomic structure of CRISPR from organism to organism. Since this mode of immunity is adaptive and takes on the spacers of phages that have previously invaded, that alone accounts for differences amongst species of bacteria and archaea.

In June 2011, many recognizable researchers in the emerging field of CRISPR, like Horvath, Koonin, van der Oost and Chaptentier, collaborated on a paper to propose a new classification for the CRISPR-Cas systems.[1] This paper rejected the system that had been in use since 2005 by Haft *et al*. and proposed a "polythetic" approach, which combined comparative genomics and structural analysis. It divided the systems into three major types and ten subtypes. Figure 2.1 outlines the classification system they proposed and the genes present in each.

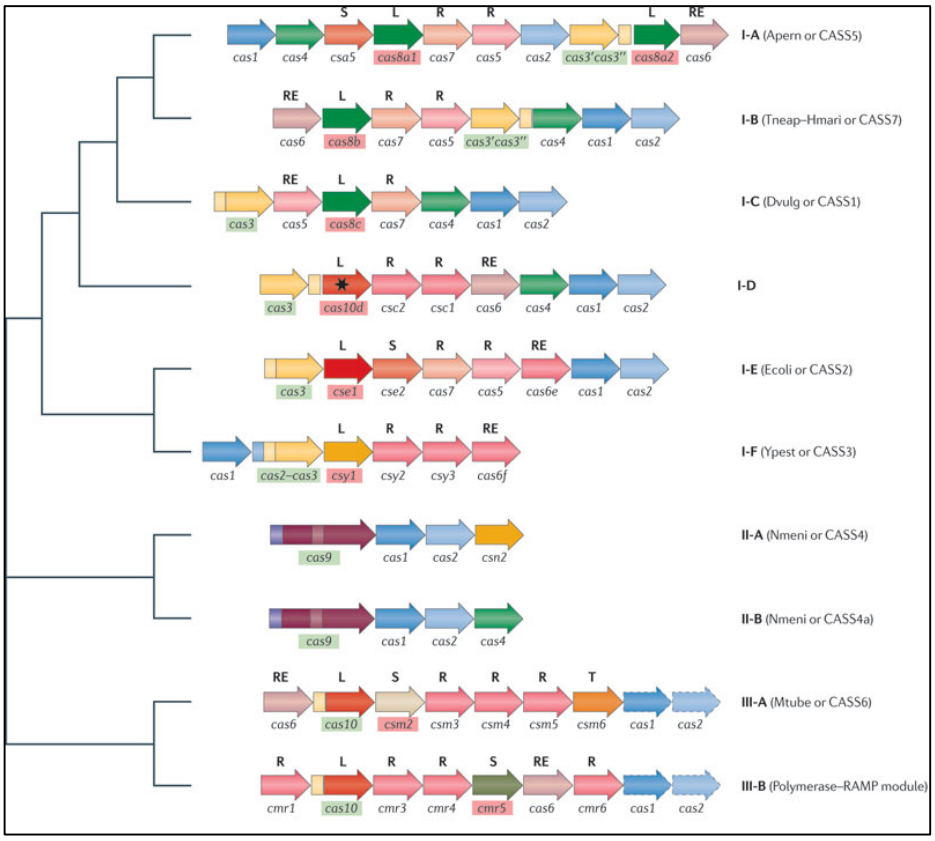


Figure 2.1: The 2011 classification of CRISPR-Cas system as proposed by Makarova *et al*.[1] This classification consisted of three major types and ten subtypes.

Revised Classification System

In November of 2015, this same core group from the 2011 paper, along with additional researchers ranging from Germany to Denmark to St. Andrews, revised their classification system to help reconcile the new discoveries being made with the growing complexity of CRISPR systems. They attempted to classify the ever changing CRISPR-cas loci by "combining the analysis of signature protein families and features of cas loci that unambiguously partitions most loci into distinct classes, types and subtypes."[2] The *cas* genes can be classified by their role in each of the three major stages in the mechanism of CRISPR: adaptation, expression and interference. For this adapted classification system though, they kept the original approach the same but used a two-

step mode of identification. The classification is based on signature genes that are unique to a particular set of CRISPR loci types and subsequent subtypes. The two-step approach involves first identifying all the cas genes in the loci, and then determining the signature ones afterwards, as opposed to only looking for the signature genes. The new classification yielded five types and sixteen subtypes, as outlined in Figure 2.2.

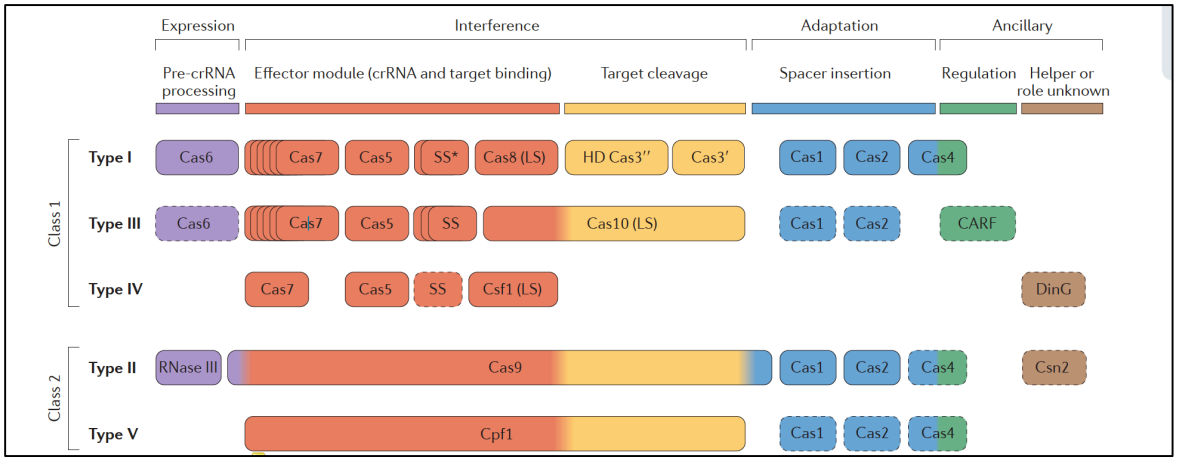


Figure 2.2: Summary of the 2015 classification system drawn with the main genes present and characterized into their primary function.[2]

Class 1 Systems

Class 1 systems are characterized as such because they use a multisubunit crRNA-effector complex.

Type 1 Systems

This is the most common type of CRISPR system found in both archaea and bacteria. The gene that defines the type I systems is the *cas3* gene. The Cas3 protein degrades single stranded DNA, shows helicase activity in the presence of ATP, and is the nuclease unit of the multi-subunit effector complex. Cas3 contains a histidine aspartate region (HD) and a metal ion binding site that contribute to its nuclease properties. The

ATP dependent helicase activity uses a ssDNA guide and is thought to take part in the adaptation of the target genome and/or the interference step.[3] Figure 2.3 is the crystal structure of the Cas3 HD domain as resolved by Mulepati and Bailey in 2011.[4] Type I systems bring about the interference of the invading DNA with the *Cascade* complex, which is a large multi-subunit protein containing the crRNA and active sites required to identify and cleave invading DNA.[5]

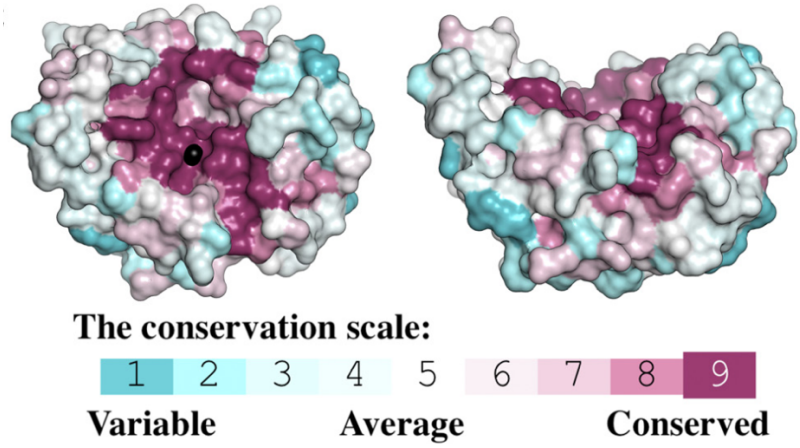


Figure 2.3: HD Domain of the *T. thermophilus* Cas3 protein with the conserved regions shaded in dark maroon.[4]

The Type I systems are divided into seven subtypes ranging from I-A to I-F and I-U, with I-U being the uncharacterized subtype because of unknown systems.[2] Subtypes I-A and I-B are encoded by two or more clustered operons while subtypes I-C to I-F are all encoded by a single operon.[2] Type I-B is the most abundant CRISPR-cas system in both archaea and bacteria. Each subtype in the Type I system differs in the way the operon is organized and the way the other *cas* genes in the system are organized. For example, the I-C systems do not have a *cas6* gene, the I-E and I-F systems do not have a *cas4* gene, and in the I-F system, the *cas2* and *3* genes are fused together. The I-U

subtypes lack a stand-alone *cas6* gene and are highly variable.[6] Makarova *et al*. did not just outline the classification, but they also showed the evolutionary ties the subtypes had to each other. Figure 2.4 outlines the relationship between the Type I systems as well as the *cas* genes present in each.

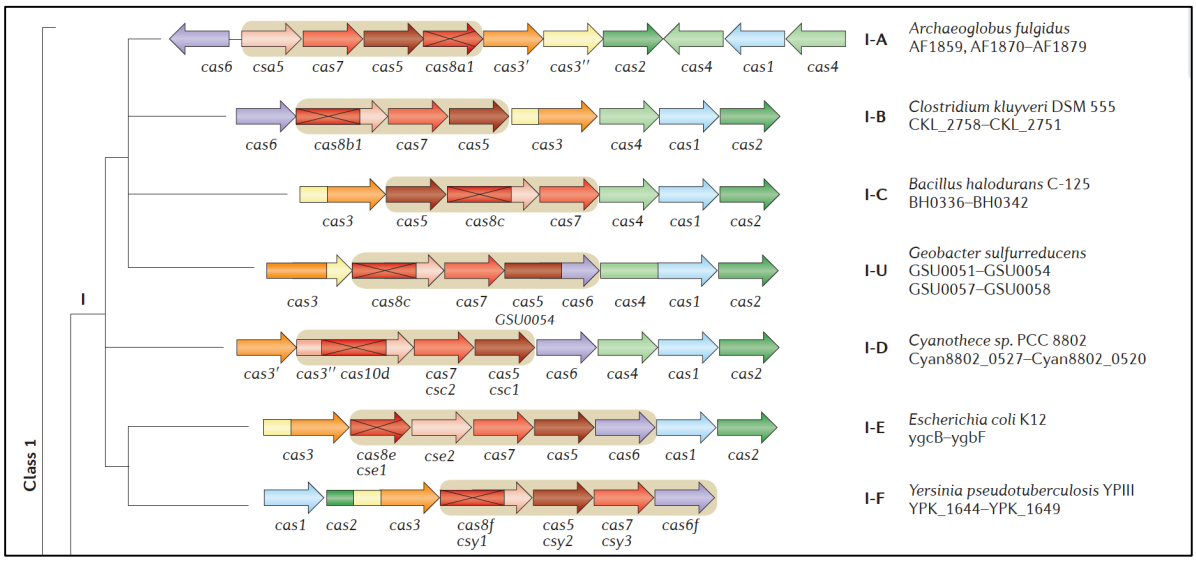


Figure 2.4: Structures and organizations of the cas genes in the Type I subtypes. The genes that encode the interference components are highlighted with a brown background. *Cas8* encodes for the large subunit of the effector complex but doesn't contain the active nuclease activity site, hence the cross through the gene. I-A and I-C are derivatives of I-B.[2]

Type III Systems

Type III systems are characterized by the signature *cas10* gene, which encodes a multi-domain protein containing a variant of a RNA recognition motif. The Cas10 protein forms the largest subunit of the effector complex of type III systems and generally contains the HD nuclease domain. Cas10 is very similar to the Cas8 (Cse1) protein of the Type I system and they appear homologous, except for the fact that Cas10 is enzymatically active.[7] Figure 2.5 is the ribbon structure of the Cmr2 protein, which at the

time of the paper[8], was only recently renamed Cas10. Previously, the type III systems were divided into two subtypes that differed on the genes that encoded the smaller subunits; III-A and III-B both were found to target RNA and DNA.[1] The recently added subtypes, III-C and III-D, varied so much from subtype A and B respectively that they could be counted as separate subtypes. Figure 2.6 shows the division of the subtypes and the genes present in each. III-C inactivates the cyclase-like domain of the Cas10 protein while the III-D loci encode a Cas10 without a HD domain. Both of these subtypes additionally lack *cas1* and *cas2* genes, which usually assist with the adaptation step.[2]

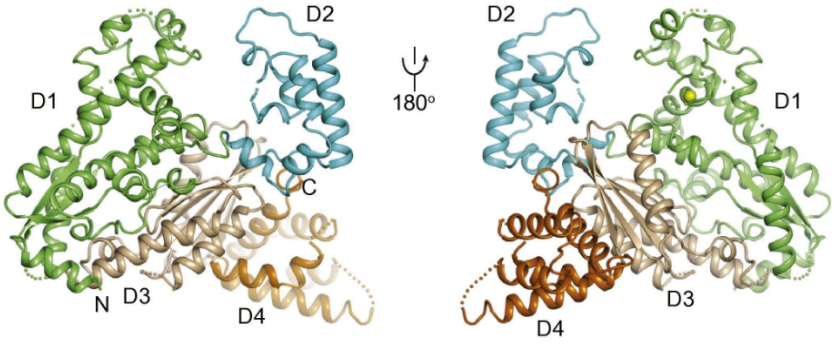


Figure 2.5: Ribbon structure of Cas10. The zinc ion is represented by a yellow sphere.[8]

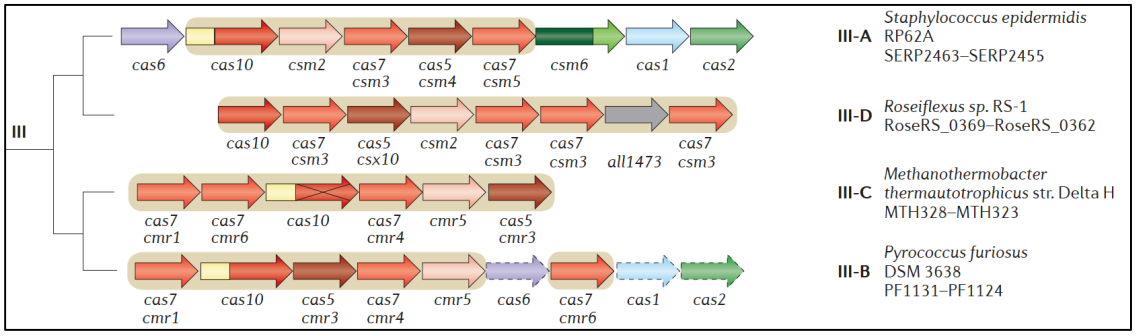


Figure 2.6: The relation of the four subtypes of a Type III CRISPR system with the interference complex genes highlighted in brown. The dashed lines for *cas1*, *cas2* and *cas6* indicate the rarity of them in III-B systems. The cross through *cas10* in III-C indicates the inactivation of the catalytic site for that subtype.[2]

Type IV Systems

Some bacteria, like the *A. ferrooxidans,* have an uncharacterized system that is labelled as the "putative type IV" CRISPR-cas system; although the cas genes that are present are not in proximity to CRISPR loci and would more appropriately be called *cas* systems.[2] They have a very minimalistic structure in their gene organization as well as in their protein effector complex. The complex has a partially degraded large subunit, and varying versions of the small subunit. Because of its high variance from types I and III, it justifies having its own subtype, but there is still much unknown about[1] (Figure 2.7).

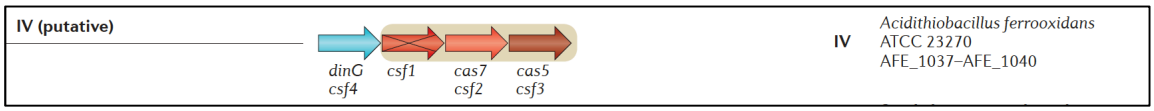


Figure 2.7: The genes constituting a type IV system. The effector complex genes are shaded in brown.[2]

Class 2 Systems

Class 2 CRISPR systems are characterized by a single subunit effector molecule, as opposed to the multi-subunit molecules of the Class 1 systems.

Type II Systems

Type II systems are the frequently utilized version of CRISPR/Cas systems, and are the type that are being explored for genetic engineering purposes. The signature gene for type II systems is *cas9,* which is transcribed into a large, multi-domain protein. Cas9, when paired with the tracrRNA:crRNA guiding complex, targets and cleaves invading DNA without the need for other components. Aside from the *cas9* gene, type II systems all also have *cas1* and *cas2,* both of which were found to be essential for the adaptation of

the foreign molecule.[2] Interestingly, the research paper that proved that *cas1* and *cas2* were required for adaptation[9] also discovered that the adaptation of new spacer units was also dependent on the presence of *cas9*. This meant that Cas9 was not simply a nuclease that cleaved the invading DNA, but that is also played a larger role earlier in CRISPR immunity.

Up until very recently, type II systems were thought to have three subtypes, two of which were part of the original classification system.[1] Figure 2.8 outlines these three well known subtypes and their differences. Type II-A systems have a *csn2* gene which is also vital to the adaptation phase[9] while II-B subtypes have a *cas4* gene in the place of the *csn2*.[2] Subtype II-C on the other hand has only three protein-coding genes and is the most common type found in bacteria. Subtypes II-C and II-A are the most abundant of the subtypes. The structure and mechanism of Cas9 will be covered in the next chapter.

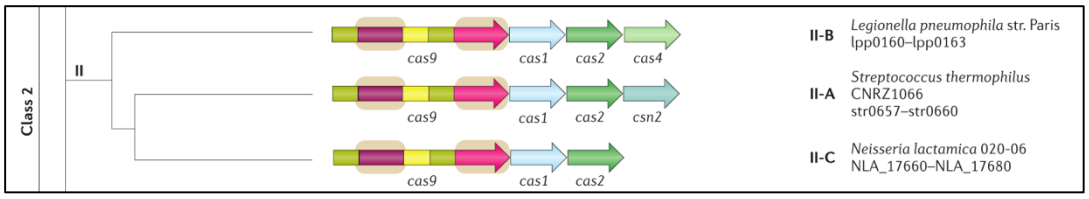


Figure 2.8: Type II subtypes with the nuclease active parts of *cas9* highlighted in brown. Subtype II-A originates from within II-C, but the minimalism and commonality of II-C allows it to be retained as a separate subtype.[2]

Type II systems were always thought to be exclusive to bacterial genomes. However, a research paper published in February of 2017 discovered previously unknown class 2 CRISPR systems within naturally occurring microbial systems and found a Cas9 system in archaea present in groundwater and sediment.[10] The archaeal Cas9 system found is thought to be a fusion of both type II-B and II-C systems. Figure 2.9 shows the construction of this recently discovered Cas9 system.

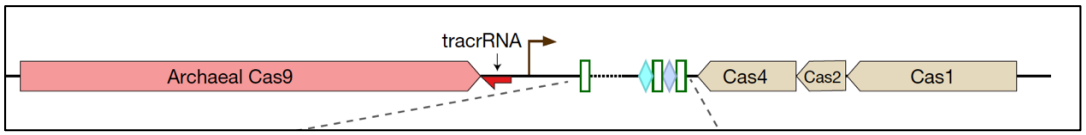


Figure 2.9: Array of CRISPR-Cas9 system found in nanoarchaea. The white boxes indicate the repeats and the diamonds indicate the spacers.[10]

Type V Systems and the Putative Type VI

Similar to the type IV systems in class 1, type V systems are putative and contain an interesting and unusual gene named *cpf1*.[2] Cpf1 is a homologue to Cas9 and is also a large protein with a RuvC-like nuclease domain.[11] However, though it lacks a HNH domain, which is characteristic of Cas9 proteins, it is still considered a class 2 system because it is a single subunit effector complex.[12] This gene is generally not found closely associated with the CRISPR-Cas loci and is sometimes encoded as stand-alone genes.[2] Cpf1 is also distinct from Cas9 in that is has only one nuclease domain and uses a single crRNA guide to identify and cleave invading DNA, thus being of interest to researchers looking for systems to manipulate for genetic engineering.[13,14] Figure 2.10 shows the architecture of one such type V system.

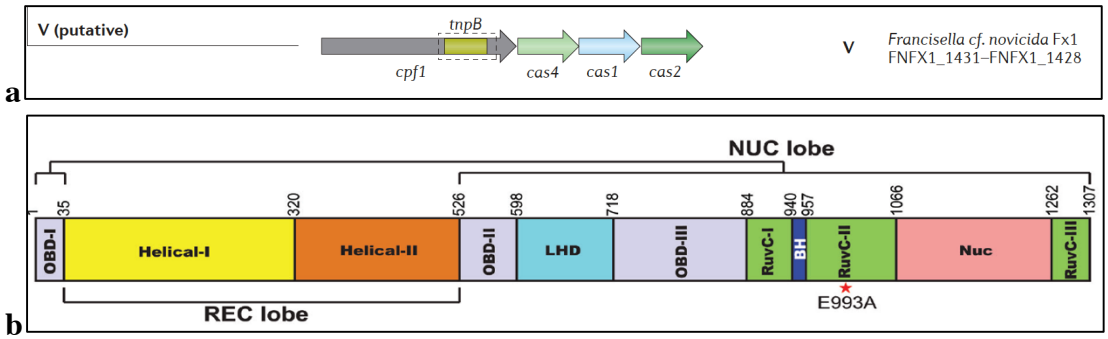


Figure 2.10: a) Type V system architecture. The lime green part of the *cpf1* gene indicates the RuvC-like nuclease while the gray parts were unclassified at the time[2]; b) The domain organization of the *cpf1* gene that was mapped after the crystal structure was elucidated in 2016 by Gao *et al*.[15]

Shortly after the publication of Marakova's classification system, Zhang *et al*. released a paper in which they used computational tools to classify and characterize two additional subtypes of type V systems, and add a potential type VI system.[16] They found single subunit effector complexes C2c1 and C2c3, both of which are related to Cpf1. The C2c2 system however, contained a predicted RNase domain and was better placed in a different type altogether—the putative type VI system. Following this, Zhang's research group published a paper online in January of 2017 where, using the same computational analysis tools as before, they were able to tentatively classify five new subgroups in type V and three new subtypes in type VI.[17] The new type V systems were classified as subgroup V-U1 through VU-5 (the U being because they are not confirmed as type V sequences yet). The effector complexes all contain RuvC nucleases. They type VI sequences are classified at VI-B1, VI-B2 and VI-C, and they all contain RNAase domains. Figure 2.11 gives an overview of the updated divisions of the type V and VI systems.

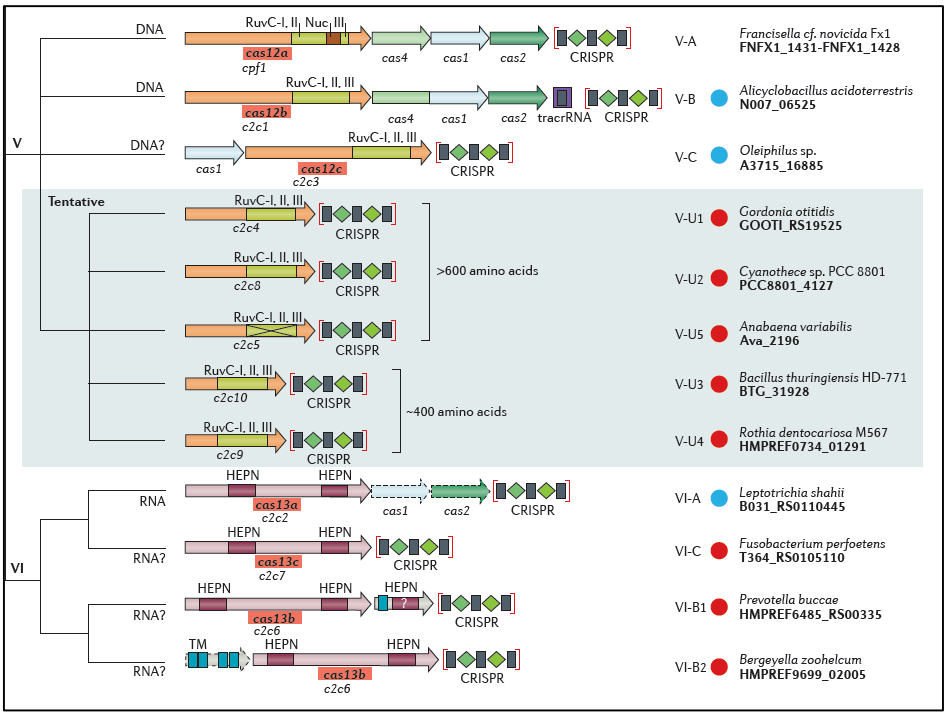


Figure 2.11: The division and relation of the newly discovered type V and VI CRISPR systems. The red circles indicate any tentative systems discovered in the 2017 paper, while the blue circles indicate the systems found in the 2015 paper.[17]

The Unclassifiable

Because of the rapid evolutionary nature of the CRISPR-cas systems, there are constantly new types of systems emerging that have not even been detected, let alone classified. It is currently not possible to account for all the variants of CRISPR-Cas systems in bacteria and archaea.[2] Initially, it was thought that *cas1* was the signature gene for all CRISPR-cas systems because it was the most highly conserved gene.[1] However, it was later found that more than 80 genomes had the CRISPR mechanism of immunity but lacked *cas1* genes.[2]

Early in 2017, the same research group that discovered the Cas9 in archaea also reported two previously unclassified and unidentified types of CRISPR-cas systems.[10] It

is too preliminary to classify their type and subtype, but Burstien *et al*. were able to determine that CRISPR-CasX is a class 2 system as it contains *cas1*, *cas2*, *cas4* and the unidentified, large CasX protein. They also discovered an extremely compact CRISPR system consisting of only a *cas1* and *casY* gene. These two systems were only discovered in the metagenomic sequencing of DNA from natural microbial communities, but are viable types of CRISPR arrays. They are very unique in how compact the systems are, but CasX is especially interesting because it is the smallest Cas protein that has been found to date.[18] This kind of discovery proves that there is still a vast amount of untapped CRISPR systems and in the years to come, the classification system will need to be modified again to include the discovery of systems like these.

[1] Makarova, K. S.; Haft, D. H.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Wolf, Y. I.; Yakunin, A. F.; *et al*. Evolution and Classification of the CRISPR-Cas Systems. *Nat. Rev. Microbiol.* **2011**, *9* (6), 467–477.

[2] Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; *et al*. An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nat. Rev. Microbiol.* **2015**, *13* (11), 722–736.

[3]Sinkunas, T.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas3 Is a Single-stranded DNA Nuclease and ATP-dependent Helicase in the CRISPR/Cas Immune System. *EMBO J.* **2011**, *30* (7), 1335–1342.

[4] Mulepati, S.; Bailey, S. Structural and Biochemical Analysis of Nuclease Domain of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Associated Protein 3 (Cas3). *J Biol. Chem.* **2011**, *286* (36), 31896–31903.

[5] Van, E.; Jackson, R. N.; Carter, J.; Golden, S. M.; Bailey, S.; Wiedenheft, B. Mechanism of CRISPR-RNA Guided Recognition of DNA Targets in *Escherichia coli*. *Nucleic Acids Res.* **2015**, *43* (17), 8381–8391.

[6] Makarova, K. S.; Wolf, Y. I.; *et al*. An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nat. Rev. Microbiol.* **2015**, *13* (Supl. S4), 722–736.

[7] Makarova, K. S.; Aravind, L.; Wolf, Y. I.; Koonin, E. V. Unification of Cas Protein Families and a Simple Scenario for the Origin and Evolution of CRISPR-Cas Systems. *Biol. Direct* **2011**, *6*, 38.

[8] Zhu, X.; Ye, K. Crystal Structure of Cmr2 Suggests a Nucleotide Cyclase-Related Enzyme in Type III CRISPR-Cas Systems. *FEBS Lett.* **2012**, *586* (6), 939–945.

[9] Wei, Y.; Terns, R. M.; Terns, M. P. Cas9 Function and Host Genome Sampling in Type II-A CRISPR–Cas Adaptation. *Genes Dev.* **2015**, *29* (4), 356–361.

[10] Burstein, D.; Harrington, L. B.; Strutt, S. C.; Probst, A. J.; Anantharaman, K.; Thomas, B. C.; Doudna, J. A.; Banfield, J. F. New CRISPR–Cas Systems from Uncultivated Microbes. *Nature* **2017**, *542* (7640), 237–241.

[11] Schunder, E.; Rydzewski, K.; Grunow, R.; Heuner, K. First Indication for a Functional CRISPR/Cas System in *Francisella tularensis*. *Int. J. Med. Microbiol.* **2013**, *303* (2), 51–60.

[12] Makarova, K. S.; Koonin, E. V. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol. Biol.* **2015**, *1311*, 47–75.

[13] Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; *et al*. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163* (3), 759–771.

[14] Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; DeGennaro, E. M.; Winblad, N.; Choudhury, S. R.; Abudayyeh, O. O.; Gootenberg, J. S.; *et al*. Multiplex Gene Editing by CRISPR-Cpf1 Using a Single crRNA Array. *Nat. Biotech.* **2017**, *35* (1), 31–34.

[15] Gao, P.; Yang, H.; Rajashankar, K. R.; Huang, Z.; Patel, D. J. Type V CRISPR-Cas Cpf1 Endonuclease Employs a Unique Mechanism for crRNA-Mediated Target DNA Recognition. *Cell Res.* **2016**, *26* (8), 901–913.

[16] Shmakov, S.; Abudayyeh, O. O.; Makarova, K. S.; Wolf, Y. I.; Gootenberg, J. S.; Semenova, E.; Minakhin, L.; Joung, J.; Konermann, S.; Severinov, K.; *et al*. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell* **2015**, *60* (3), 385–397.

[17] Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; *et al*. Diversity and Evolution of Class 2 CRISPR-Cas Systems. *Nat Rev Micro* **2017**, *15* (3), 169–182.

[18] Cross, R. Expanding CRISPR Toolkit May Render Patent Fight Moot | Chemical & Engineering News. *C&EN*, 2016.

CHAPTER THREE

The Mechanism of Cas9 and Cascade CRISPR Systems

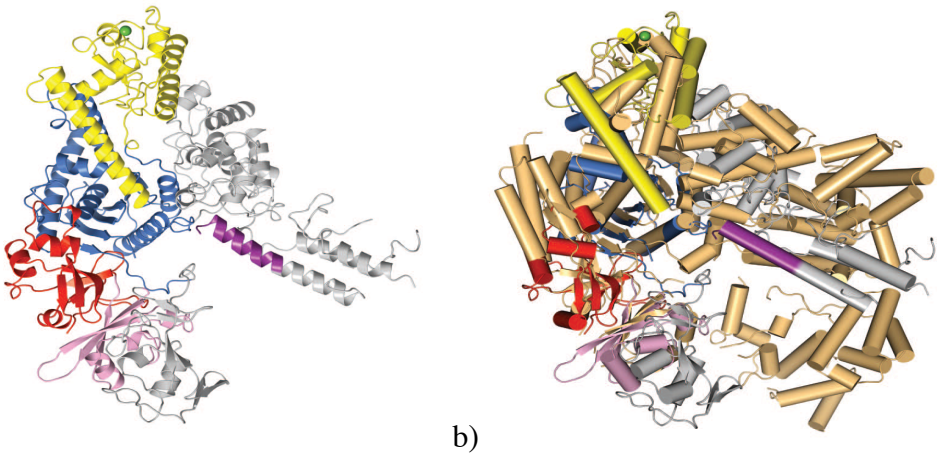
Type II systems and the Cas9 Protein

The most frequently researched and tested CRISPR systems are the Type II systems that are present primarily in bacteria (but have very recently been found in some archaea as well[1]). The simplistic Cas9 effector protein allows for a good way to investigate the function of the CRISPR system and it is the protein currently being used for most genome editing and research.

The function of CRISPR systems can be divided into three main stages; adaptation, expression, and interference. The palindromic repeats that are acquired from external genomes are adapted into the host's DNA. The Cas proteins aid in acquiring the external DNA and in bringing about the immune response by targeting DNA. For Type II systems, the Cas9 protein is by far the most important to the functioning of the system and to the mechanism of DNA cleavage. This is why understanding the structure of Cas9 is vital to understanding of the mechanism behind Type II-CRISPR immunity.

Over the course of 2014 and early 2015, multiple papers emerged with crystal structures of Cas9 in various stages and in various bound and unbound states.[2] Together, these structures helped elucidate the Cas9 endonuclease mechanism of cleavage of invading DNA. Cas9 is a large protein that varies greatly in size from organism to organism, but overall shows a bilobed architecture and contains a highly conserved catalytic domain. Cas9 proteins appear to adopt a catalytically inactive conformation

when in the apo form and changes conformation to be catalytically active for DNA cleavage.[3] The conserved catalytic core of Cas9, as shown in Figure 3.1, contains both a HNH domain ("HNH" describing the amino acid residues in the active site) that cleaves the complementary DNA strand, and a RuvC nuclease domain that cleaves the non-complimentary strand, yielding the double stranded break.[3] The HNH domain generally contains two Histidine residues and an Asparagine. Cas9 additionally has catalytic residues similar to those in HNH nucleases; Asp839, His840, and Asn863, with the histidine residue being critical to the cleavage.[4]



a) b)

Figure 3.1: (a) Structure of the conserved catalytic core from *A. naeslundii* Cas9 (AnaCas9) in ribbon form. The RuvC domains are in blue, HNH in yellow, and the α-helical lobe in gray. (b) The AnaCas9 superimposed over *S. pyogenus* Cas9 (SpyCas9), which is colored in orange.[3]

The two lobes of the Cas9 protein are the Nuclease lobe (NUC) and the recognition lobe (REC). The NUC lobe contains the RuvC, HNH and PAM-interacting (PI) domains, all of which are crucial to the positioning and subsequent cleavage of any DNA strand. Nishimasu *et al*. in 2014 resolved the crystal structure of Cas9 while bound

to the gRNA as well as the target DNA, which allowed them to make conclusive statements on the Cas9 mechanism.[4] Figure 3.2 is the crystal structure from Nishimasu *et al*., while Figure 3.3 is the layout of the *cas9* gene with the domains labeled.

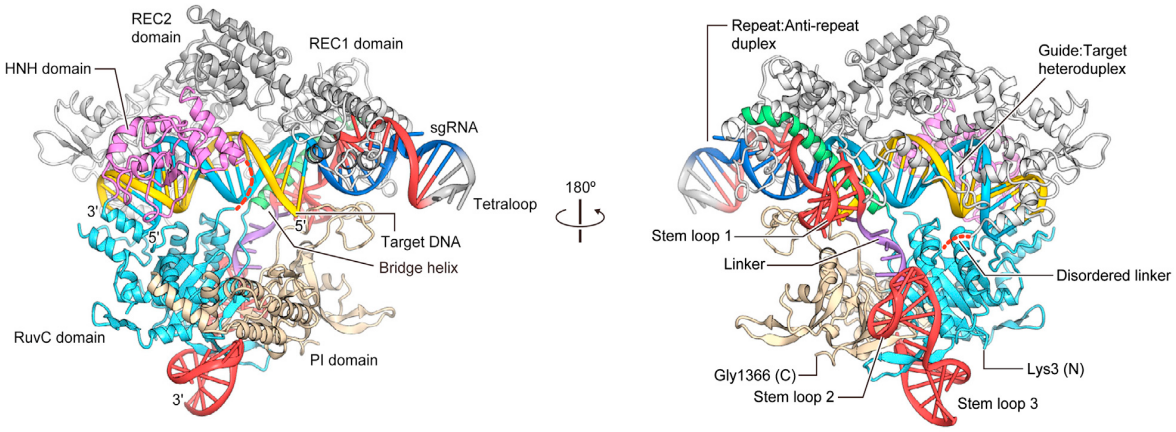


Figure 3.2: Ribbon structure of SpyCas9 (*S. pyogenes*) with the labelled domains.[4]

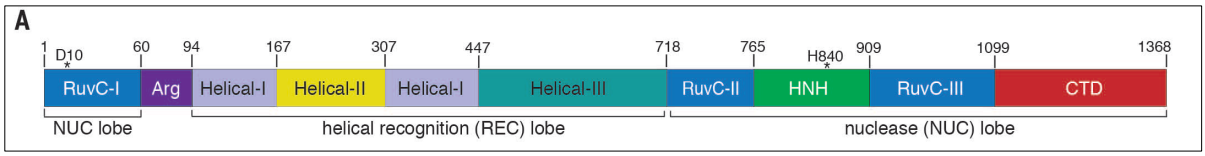


Figure 3.3: Arrangement of domains in the *cas9* gene as found in SpyCas9.[5]

Figure 3.4 from Jiang's 2015 study of the crystal structure of the Cas9 of *S. pyogenes* is a depiction of their overall proposed mechanism. When the target DNA enters the Cas9, the guide RNA recognizes the PAM sequence, unwinds a portion of the target DNA and forms a heteroduplex. The Cas9 protein forces this heteroduplex to form a T-shape[4] and bends it 30 degrees, thus providing the structural distortion needed for R-loop formation.[6] This R-loop that is formed inside the Cas9 catalytic site is what allows for the DNA to be nicked by the RuvC and HNH domains three base pairs upstream of the PAM site.[6] The formation of the heteroduplex and the bending of the DNA allows the HNH domain to shift and cleave the strand complementary to the guide RNA, while

the RuvC cleaves the non-complementary strand.[4] This is the conformational change that is required to make Cas9 catalytically active and accommodate the DNA and guide RNA.[5]

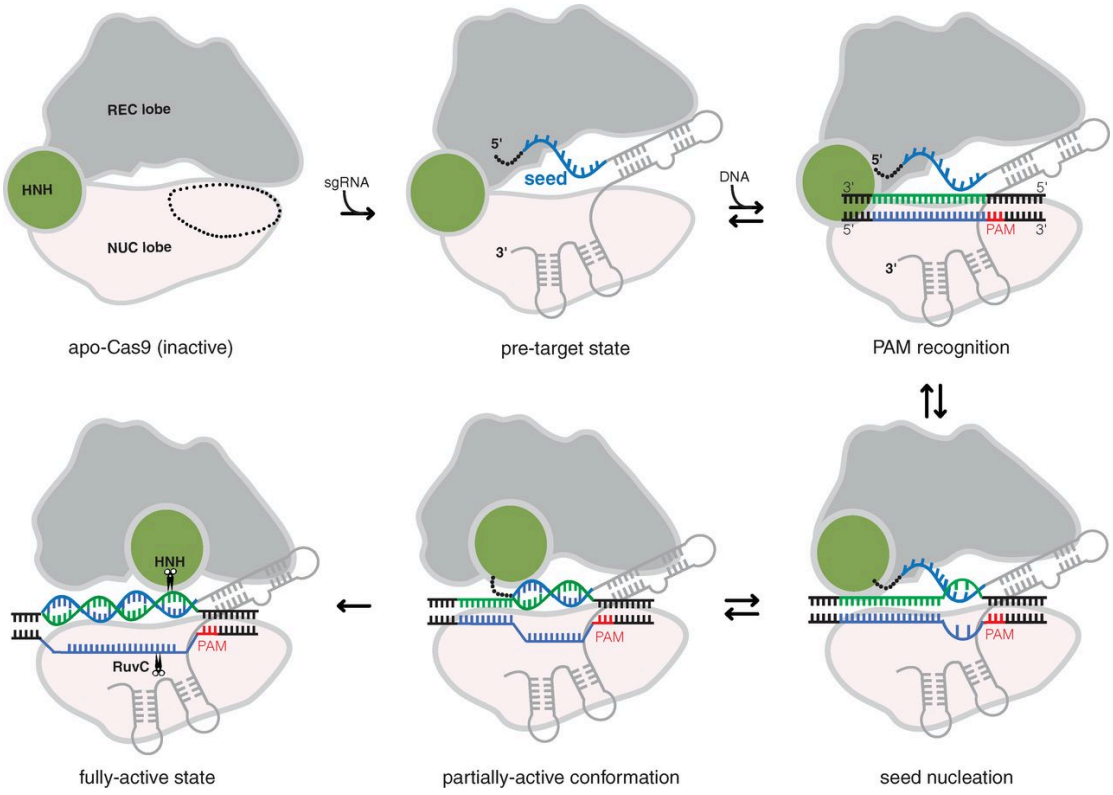


Figure 3.4: Cartoon depiction of the proposed mechanism of action of Cas9. The first conformational change occurs when the guide RNA binds and the second major change occurs when the base pairing between the seed sequence and target DNA cause the HNH domain to move into place for cleavage.[5]

The mechanism of action for CRISPR systems begin with the acquisition of spacers. Cas1 and Cas2 are the only proteins that are conserved across a majority of the various types and subtypes of CRISPR systems and are vital for the adaption of foreign spacers.[7] Cas1 and Cas2 work together and bind to invading DNA at positions determined by the presence of the PAM (Protospacer Adjacent Motif) site.[8] Cas1 has nuclease activity and the Cas1-Cas2 complex selects a very precise 30-35 nucleotide long

protospacer from the foreign bacteria. The general consensus is that the host cell picks these dsDNA pieces from degraded fragments arising out of DNA repair mechanisms.[9,10] When the DNA of the phage is damaged and is undergoing repair, Cas1-Cas2 can bind to free 3' –OHs and act as a molecular ruler to pick out exactly ~30 bp of a dsDNA protospacer.[9, 11] It has also been suggested that polymerase activity is involved and the DNA is replicated instead of taken directly from the virus, but there has not been much evidence supporting this.[12] Additionally, it has been found that Cas9 is vital to the incorporation of the spacers, and though it is not yet clear how, it is known that the nuclease activity does not play a role in the adaptation.[13]

Once the foreign protospacer DNA has been acquired by the Cas1-Cas2 complex, it then needs to be incorporated into the host's genome. New spacers are found to be incorporated upstream of the first repeat near the 5' end and the 200-500 bp sequence preceding it, which contains the promotor, appears to play a role in guiding the addition of the new spacer.[12,14] Nuñez *et al*. proposed a two-step mechanism for the incorporation of the spacer into the CRISPR array. Figure 3.5 summarizes this mechanism and shows how the 3' –OH on the protospacer attacks a 3' –OH on the host's DNA.

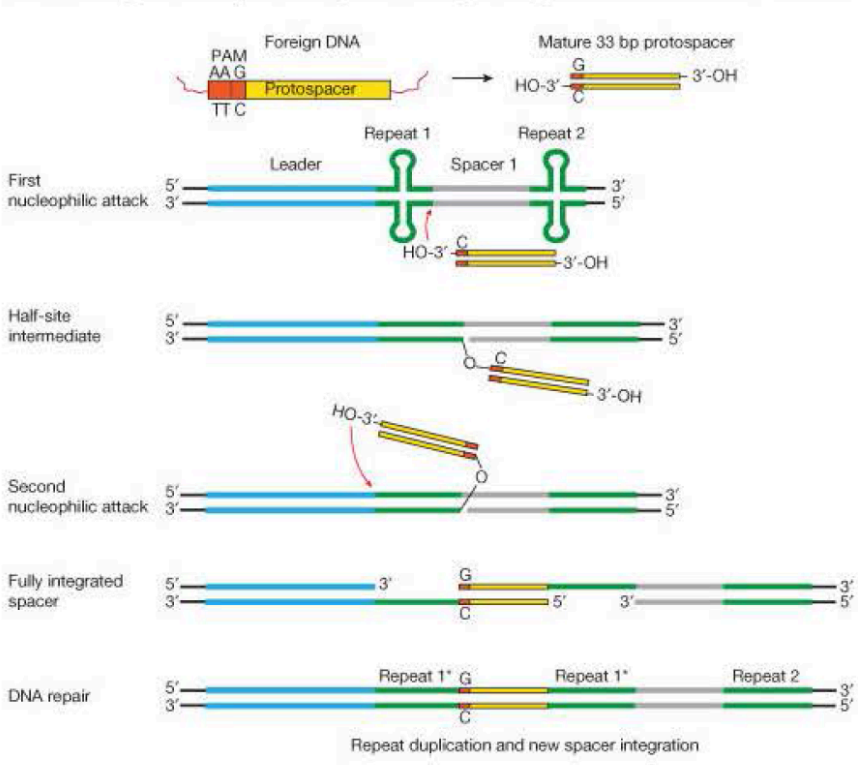


Figure 3.5: Model depicting the incorporation of the protospacer into the CRISPR array showing the 3' –OH attack.[14]

With the incorporation of a new spacer comes the acquired immunity. The next time the same phage attacks, the CRISPR pathway is activated and sends out the Cas9 protein with the associated guide RNA. The guide RNA is made up of portions of the CRIPSR array that are not translated and consist of a *trans*-activating CRISPR RNA (tracrRNA) and the spacer/repeat derived sequence of CRISPR-RNA (crRNA).[15] When the time for an immune response comes, the tracrRNA and its extensive secondary structure allow for the maturation of inactive pre-crRNA to active crRNA. This process forms the folded dsRNA complex that is loaded into the Cas9 protein and makes use of RNAse III–the only non-CRISPR related protein in the whole process.[15,16] RNAse III functions to trim the 3' end of the pre-crRNA thus preparing the mature 3' –OH end of crRNA.[17] The formation of mature effector complexes does not occur in the absence of RNAase III, tracrRNA or Cas9, indicating Cas9 also has a role to play in the maturation

of the guide RNA complex.[15] Figure 3.6 depicts a simplified model of the pathway for the maturation of the tracrRNA:crRNA complex. Following this maturation of the complex, the Cas9 protein can go on to identify and cleave the foreign DNA with the mechanisms discussed above.

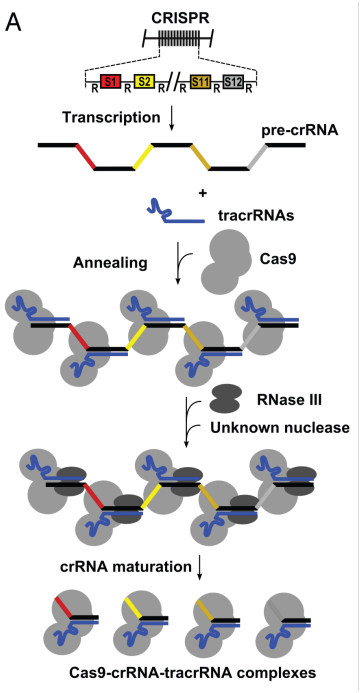


Figure 3.6: A model of the mechanism of CRISPR RNA maturation and effector complex formation.[17]

Type I systems and the Cascade Complex

The other highly studied CRISPR systems are the Class 1 Type I systems that are highly prevalent in bacteria and archaea. They are characterized by the multi-subunit complex, employed for all aspects of CRISPR function, as well as by the *cas3* gene.[7] The focus of this section will be the Cascade complex specifically in *E. coli*. The multi-subunit effector complexes in Type-I systems differ from subtype to subtype in the

number of subunits and types of *cas* genes that are associated with it. The one commonality between all of the subtypes is the recruitment of the Cas3 protein to bring about the final degradation of the DNA.[7]

The cascade complex in *E. coli* consists of a sequence of eight genes that code for the subunits of the effector complex as well as the other proteins involved in acquiring and degrading the DNA.[18] Cas3 is the protein recruited for the degradation of the foreign DNA, Cas1 and Cas2 are required for spacer acquisition[19] and CasABCDE form the effector complex. Figure 3.7 illustrates the layout of the genes for *E. coli* and the proportions in which the Cascade genes are present. There is one subunit each of CasA, CasD and CasE, two subunits of CasB subunits and six of CasC. The six CasC proteins form the backbone of the complex and are arranged as a helical stack while the crRNA forms the spine and lies against the backbone in a horseshoe shape. CasE sits at 3' end of the mature crRNA as the tail while CasA (Cse1) acts as the head.[20]

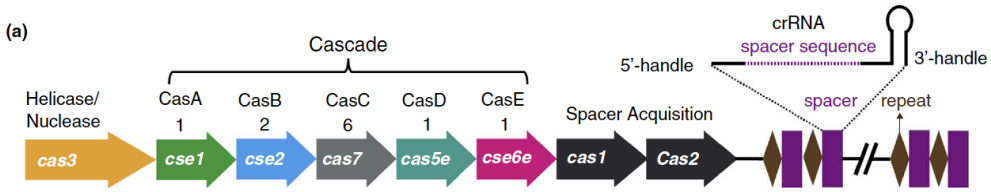


Figure 3.7: Gene sequence of the Type I *E. coli* CRISPR system. The Cascade complex is labelled from CasA to CasE, but the formal names are written within the arrows. The numbers above the genes correspond to the stoichiometry of each subunit.[18]

CasE is an endoribonuclease that is vital to the pre-crRNA cleavage and maturation.[21] The crRNA and DNA are able to bind to the complex well because of lysine rich positive areas at each end of Cascade. The positively charged residues allow for non-sequence-specific binding and for DNA to bind even if the correct PAM or protospacer sequences are not present.[22] The entire complex undergoes a conformational

change upon the binding of the target DNA. It is thought that this conformational change is what allows for the Cas3 protein to be recruited for the degradation of the exposed DNA.[20] Figure 3.8 shows the various depictions of the structure of Cascade while Figure 3.9 is a model of the function of cascade.

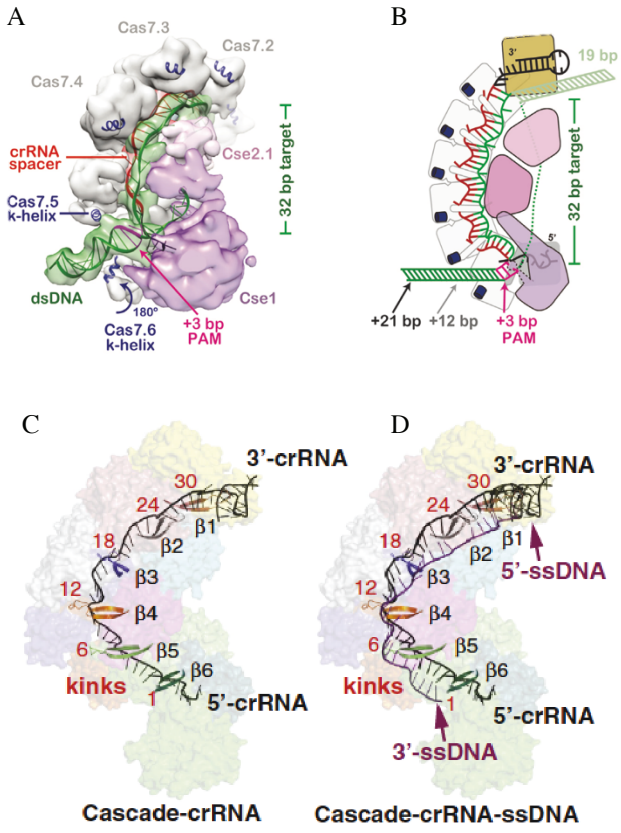


Figure 3.8: A) 3D model of Cascade bound to a dsDNA with the lysine helices indicated by blue ribbons.[22] B) A schematic of the binding of dsDNA to the cascade complex along with the number of base pairs required to reach the lysine sequences at the end of the complex.[22] C and D) Two depictions of the internal binding and positioning of both crRNA as well as crRNA-ssDNA.[18]

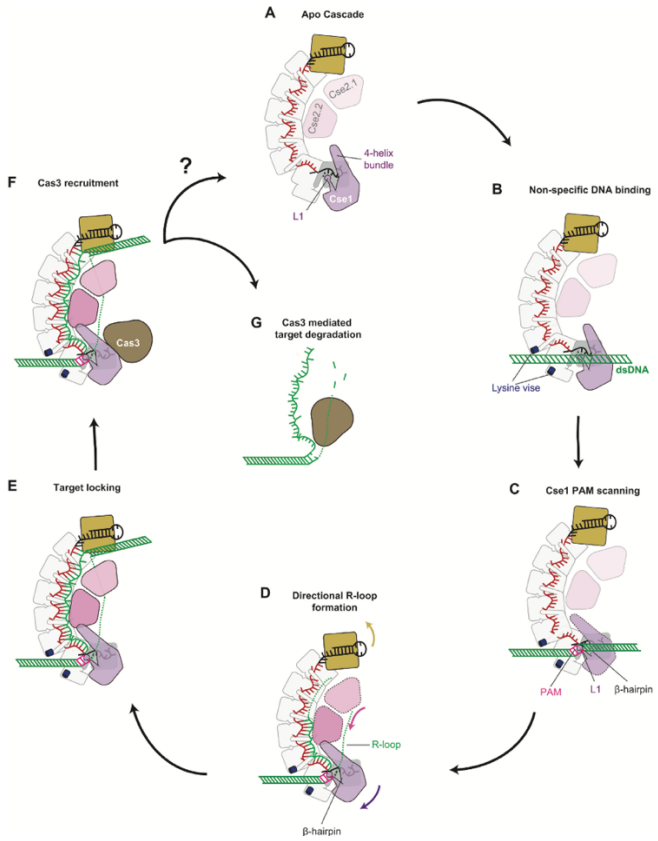


Figure 3.9: Target recognition and binding by Cascade. Step F shows the recruitment of Cas3 for degradation, but it is not known how the complex is regenerated.[22]

Despite there being little similarity between Type III and Type I Cas sequences, there is commonality in the arrangement of the effector complexes. Type III complexes have a similar construction of the smaller subunits but a large subunit that is the Cas10 protein, instead of the Cas8 (Cse1) found in Type I systems. The major difference between the large subunits is that Cas10 is catalytically active in Type III systems and therefore doesn't require Cas3 to degrade the foreign DNA/RNA.[23] Additionally, Type III systems do not require *cas1* or *cas2* to acquire new spacers; instead, it is thought they use *cas6*.[24] Type III systems, while fewer in number, are more diverse than Type I in their composition and organization, and there is still much more to be learned about their mechanism of action.

[1] Burstein, D.; Harrington, L. B.; Strutt, S. C.; Probst, A. J.; Anantharaman, K.; Thomas, B. C.; Doudna, J. A.; Banfield, J. F. New CRISPR–Cas Systems from Uncultivated Microbes. *Nature* **2017**, *542* (7640), 237–241

[2] Wang, H.; Russa, M. L.; Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. of Biochem.* **2016**, *85* (1), 227–264.

[3] Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; *et al*. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **2014**, *343* (6176), 1247997.

[4] Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156* (5), 935–949.

[5] Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. A Cas9–guide RNA Complex Preorganized for Target DNA Recognition. *Science* **2015**, *348* (6242), 1477–1481.

[6] Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A. Structures of a CRISPR-Cas9 R-Loop Complex Primed for DNA Cleavage. *Science* **2016**, *351* (6275), 867–871.

[7] Makarova, K. S.; Haft, D. H.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Wolf, Y. I.; Yakunin, A. F.; *et al*. Evolution and Classification of the CRISPR-Cas Systems. *Nat. Rev. Microbiol.* **2011**, *9* (6), 467–477.

[8] Horvath, P.; Romero, D. A.; Coûté-Monvoisin, A.-C.; Richards, M.; Deveau, H.; Moineau, S.; Boyaval, P.; Fremaux, C.; Barrangou, R. Diversity, Activity, and Evolution of CRISPR Loci in *Streptococcus thermophilu*s. *J. Bacteriol.* **2008**, *190* (4), 1401–1412.

[9] Nuñez, J. K.; Harrington, L. B.; Kranzusch, P. J.; Engelman, A. N.; Doudna, J. A. Foreign DNA Capture during CRISPR–Cas Adaptive Immunity. *Nature* **2015**, *527* (7579), 535–538.

[10] Amitai, G.; Sorek, R. CRISPR-Cas Adaptation: Insights into the Mechanism of Action. *Nat. Rev. Microbiol.; London* **2016**, *14* (2), 67–76.

[11] Levy, A.; Goren, M. G.; Yosef, I.; Auster, O.; Manor, M.; Amitai, G.; Edgar, R.; Qimron, U.; Sorek, R. CRISPR Adaptation Biases Explain Preference for Acquisition of Foreign DNA. *Nature* **2015**, *520* (7548), 505–510.

[12] Heler, R.; Marraffini, L. A.; Bikard, D. Adapting to New Threats: The Generation of Memory by CRISPR-Cas Immune Systems. *Mol. Microbiol.* **2014**, *93* (1), 1–9.

[13] Wei, Y.; Terns, R. M.; Terns, M. P. Cas9 Function and Host Genome Sampling in Type II-A CRISPR–Cas Adaptation. *Genes Dev.* **2015**, *29* (4), 356–361.

[14] Nuñez, J. K.; Lee, A. S. Y.; Engelman, A.; Doudna, J. A. Integrase-Mediated Spacer Acquisition During CRISPR-Cas Adaptive Immunity. *Nature* **2015**, *519* (7542), 193–198.

[15] Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III. *Nature* **2011**, *471* (7340), 602–607.

[16] Marraffini, L. A. The CRISPR-Cas System of *Streptococcus pyogenes*: Function and Applications. In *Streptococcus pyogenes : Basic Biology to Clinical Manifestations*; Ferretti, J. J., Stevens, D. L., Fischetti, V. A., Eds.; University of Oklahoma Health Sciences Center: Oklahoma City (OK), 2016.

[17] Karvelis, T.; Gasiunas, G.; Miksys, A.; Barrangou, R.; Horvath, P.; Siksnys, V. crRNA and tracrRNA Guide Cas9-Mediated DNA Interference in *Streptococcus thermophilus*. *RNA Biol*. ***2013****, 10 (5), 841–851*.

[18] Jiang, F.; Doudna, J. A. The Structural Biology of CRISPR-Cas Systems. *Curr. Opin. in Struct. Biol*. **2015**, *30*, 100–111.

[19] Yosef, I.; Goren, M. G.; Qimron, U. Proteins and DNA Elements Essential for the CRISPR Adaptation Process in *Escherichia coli*. *Nucleic Acids Res*. **2012**, *40* (12), 5569–5576.

[20] Wiedenheft, B.; Lander, G. C.; Zhou, K.; Jore, M. M.; Brouns, S. J. J.; van der Oost, J.; Doudna, J. A.; Nogales, E. Structures of the RNA-Guided Surveillance Complex from a Bacterial Immune System. *Nature* **2011**, *477* (7365), 486–489.

[21] 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, D. O. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **2008**, *321* (5891), 960–964. (in previous chapter).

[22] Van, E.; Jackson, R. N.; Carter, J.; Golden, S. M.; Bailey, S.; Wiedenheft, B. Mechanism of CRISPR-RNA Guided Recognition of DNA Targets in *Escherichia coli*. *Nucleic Acids Res*. **2015**, *43* (17), 8381–8391.

[23] Makarova, K. S.; Zhang, F.; Koonin, E. V. SnapShot: Class 1 CRISPR-Cas Systems. *Cell* **2017**, *168* (5), 946–946.e1.

[24] Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; et al. An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nat. Rev. Microbiol*. **2015**, *13* (11), 722–736.

CHAPTER FOUR

Harnessing the Tool

With the publishing of Siksnys' paper in 2011, which found that a CRISPR system from one organism could function in another,[1] it was clear that the world of gene editing would be drastically transformed. The next big step was taken by Charpentier and Doudna, when they were able to prove that CRISPR/Cas9 not only functions *in vitro,* but is programmable and can be modified to cleave any sequence of DNA.[2] This discovery opened the door for a flood of research into the programmable aspects of Cas9 and instigated advances in work to perfect the technology. Once researchers like Feng Zhang[3] and George Church[4] showed that this technology could be used in humans, it was clear how valuable CRISPR would be. The ability to edit genes and genomes has been essential to our growing understanding of genes and their functions. Beyond the ability to study genes, the fully sequenced human DNA and a precise editing tool has given us the ability to manipulate genes for therapeutic purposes; the more precise the editing mechanism, the more applicable the technique can be to human genetic disorders.[5] Precise methods are especially required for any editing involving human genomes because off-target effects could have unprecedented results and could be a huge expense, only to yield unsuccessful results. The emergence of a simple and programmable system like CRISPR-Cas9 gives scientists a cheap and fast way to edit genes while minimizing off-target effects.[5,6] When researchers had figured out they could program Cas9, the task then emerged to turn it into a tool that can be used by any researcher in any lab.

From the Lab to the Market

Existing gene editing mechanisms that involve nucleases, like Zinc-Finger Nucleases (ZFN)[7] and transcription-like effector nucleases (TALENs)[8] utilize dimerization to surround the target DNA and induce a double stranded break. However, while widely used, these methods are somewhat inefficient, expensive, difficult to develop and come with uncertainty as to the strength of dimerization and precision of nicking.[6, 7] While Cas9 provides the means to easily cleave at a specific point on the target DNA, the actual "editing" process has to be manipulated afterwards by making use of existing repair mechanisms present in the cell. As Charpentier and Doudna were perfecting their system for programming Cas9 to be used in prokaryotes, Feng Zhang and his team at MIT were modeling the technology for use in eukaryotes. Their 2013 procedure paper gives, in full detail, the methodology for designing and using programed CRISPR-Cas9 to edit genes in mammalian cells[6].

Previously, in order to design a Cas9 system that will target a DNA sequence of your choice, you had to find a sequence that immediately preceded a PAM site (NGG), so that the Cas9 could identify the target.[6] However, it was then discovered that it is possible to mutate the PAM recognition site on the Cas9 protein to recognize other PAM sequences, making it is possible to relax the binding on the third base site, thus broadening the range of PAM binding.[9,10] There is still more research to be done in these areas to ensure the efficiency of the Cas9 being used. Once the target site has been identified on the DNA, online sequencing tools assist in designing the appropriate sgRNA sequence required. If edits are being made to the target DNA, these same sequencing tools can help design the altered DNA oligonucleotide that will serve as the

template for the DNA repair. The Cas9 and sgRNA are expressed in plasmids and can subsequently be delivered into mammalian cells in a variety of ways. For example, they can be delivered via a lipophilic transporter or calcium phosphate, or via a viral transport.[11]

Once the Cas9 system has resulted in a double stranded break *in vivo*, the cell's natural repair mechanisms usually come into play. It is then possible to use these repair mechanisms to bring about a desired change or addition to the DNA. Even with traditional methods of gene editing, modifications to the bases or additions were brought about by plasmid-based donor templates that shifted DNA repair in a certain direction.[6] Normally in mammalian cells, when the DNA undergoes a double stranded break, the cell brings about the repair mechanism known as nonhomologous end joining (NHEJ), where it simply chews down the DNA to blunt ends, and ligates the break.[6,12] This is a highly error prone mechanism of repair and leaves insertions or deletions, thus leading to frameshift mutations.[13] This approach can be used to inactivate genes. Cells are also capable of a second, higher fidelity pathway known as homology directed repair (HDR)[12], which involves using the DNA of the homologous chromosome as a template to fill in the missing bases. One of the stands from the homologous pair forms a Holliday junction with the damaged strand and repair polymerases use that as a template to synthesize the damaged parts.[13] Figure 4.1 outlines the mechanism for both of these repair processes. Researchers can include a template strand with the desired edit along with the Cas9 system to ensure the DNA undergoes HDR with the desired sequence of base pairs. The template can either be a double stranded DNA fragment or a constructed single stranded DNA oligonucleotide (ssODN).[6, 12]

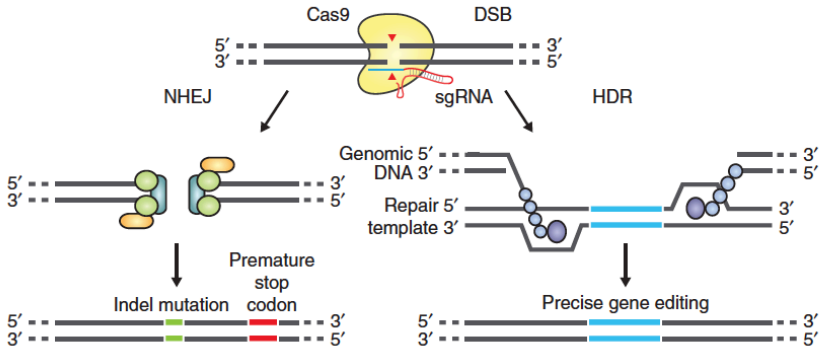


Figure 4.1: The two ways double stranded breaks can be repaired in cells, with HDR being the more accurate method, but NHEJ being the more common, yet more error prone, one.[6]

Thus, Cas9 can be engineered and used to great levels of specificity to incur mutations in DNA, or to bring about edits. Cas9 molecules can also be engineered not to cleave but only nick the DNA at one strand. These "nickase" variants are thought to allow for more precise sequence recognition and minimize off target effects.[14]

Cas9 nucleases are ultimately beneficial because they allow for cleavage of DNA at a specific position. The existing methods of gene editing like ZFNs and TALENs also allow for double stranded cleavage, but CRISPR-Cas9 has gained so much attention because it provides several major advantages over these methods. Cas9 is very easy to customize to any given target and only requires the coding of a 20-nucleotide guide sequence.[6] It is a single protein and both the sites for cleavage are within the monomer, as opposed to the dimers that are required for ZFN and TALENS.[12] It is also a relatively fast process from the coding of the sequences to the execution of the edits and is cheaper to perform.[15]

Recently, another CRISPR system—Cpf1–has been studied for its viability as a genomic engineering tool. CRISPR-Cpf1 is a Class 2 Type V system, so it is also a single subunit effector complex and uses a single RNA to guide its cleavage activity.[16] Cpf1, in some ways, is more efficient than Cas9 systems. Cpf1 systems can mature the crRNA without the help of a tracrRNA, cleave DNA with T rich PAM sequences (unlike the G-rich sequence in Cas9), and introduce a staggered double stranded break in DNA.[17] These features give it an advantage over Cas9 in accuracy and simplicity, thus introducing it to research as a viable mechanism for editing.[18]

Uses and Research So Far

Since CRISPR-Cas9 was identified as a method to edit genes efficiently, research immediately began on the viability and effectiveness of it as a therapeutic tool. Research ranging from *in vitro* studies to stem cell tests in mice and even human zygotes emerged. CRISPR is being utilized as a method to correct genetic mutations that result in diseases, as well to potentially enhance the T-cells in the human body and give them the ability to recognize antigens on cancer cells and kill them.[19] In 2014, it was shown that Cas9 was able to produce chromosome rearrangements at the loci related to lung cancer[20] and later in 2014, papers were published showing how CRISPR can be used to induce human forms of various cancers in mice and stem cells.[21,22]

Research teams in China have dived in headfirst with the T-cell modifications and are already undergoing clinical trials using CRISPR therapeutically on humans. So far, there are trials involving the T-cell modification for prostate cancer, bladder cancer and renal cancer.[19] In June of 2016, a lab in the US received permission from the NIH (National Institutes of Health) to run a clinical trial on humans. By talking T-cells from

patients with various forms of cancer and inserting a gene that would allow the T-cells to detect cancer cells, they hope to engineer T-cells to bring about cell death.[23]

Additionally, there have been countless other studies showing Cas9 being used to fix mutations or create mutations to mimic diseases in somatic cells and germ-line cells. In 2014, it was shown that somatic mutations in mice can be edited with Cas9 and the wild type variant of the hepatocyte being edited was restored[24]. In 2016, a form of human hemophilia B was induced in mice and then cured.[25] Various germ line edits have also been done on mice; one study produced mosaic mice in an attempt to cure muscular dystrophy[26], while another fixed a cataract-causing mutation that was then passed on to the mice's progeny.[27] A lab in China was able to show that effective germ-line editing in mice was possible by using CRISPR on the spermatogonial stem cells (SSC) and breeding the mutant sperm with female gametes.[28] The SSC edits produced almost no off-target effects and produced live pups, all of which had a successful transmission of the mutated gene. Human stem cells have also been used with CRISPR-Cas9 in studies as early as 2013 with corrections made to cystic fibrosis mutations.[29] Multiple tests have been conducted on pluripotent human stem cells ranging from muscular dystrophy,[30] to improving hematopoietic differentiation,[31] to hemophilia.[32]

Chinese research jumped a few steps ahead with their studies by doing two of the most surprising and ethically compromising research with CRISPR-Cas9. Firstly, in April 2015, a lab in Guangzhou, China, edited non-viable human triponuclear zygotes to eliminate certain β-thalassemia mutations.[33] However, mosaicism in the cells and unidentifiable off-target effects made it difficult to label the work a success and only underlined the need for further investigation into the clinical applications of CRISPR at

an embryonic level. Secondly, in 2015, a team of researchers bred Beagles that had their Myostatin producing gene deleted. Myostatin regulates the amount of skeletal muscle growth in dogs, so by deleting the gene, the researchers could raise puppies that had double the normal muscle mass of their breed.[34] This study also produced chimeric dogs and of the 27 that were born, only 2 showed full development with the edits.

Yours or Mine?–The Patent Battle

When it was discovered that Cas9 could be used a powerful lab tool, the race to patent the technology began. CRISPR or Cas9 itself cannot be patented since they are naturally occurring systems in bacteria and archaea, but since they do not exist naturally in mammals, the components and mechanisms to adapt Cas9 for lab use can be patented. Both Dr. Feng Zhang's lab at the Broad Institute with MIT and Charpentier and Doudna's lab at UC Berkley submitted patents for the Cas9 system of gene editing around the same time. Because of overlapping claims, the patent cases for the two had to be reviewed in court. In what became a year-long court battle, it was decided that the claims of the two patents do not overlap since Zhang's lab "were issued patents for methods for genome editing in eukaryotic (including human) cells, while UC Berkeley and collaborators applied for patents concerning CRISPR methods based on studies in cell-free systems that did not involve genome editing in eukaryotic cells"[35] (however, the ruling is now being appealed by UC Berkeley). Thus, the Broad Institute was able to keep the rights to the Cas9 system of editing in mammalian cells and they now work with Addgene, a plasmid repository, to make the CRISPR plasmids and reagent information available to the scientific community. Additionally, the leading researchers in CRISPR,

like Zhang, Church, and Liu, founded Editas Medicine in 2013. Editas is a discovery based pharmaceutical company that is using the CRISPR technologies developed at MIT to therapeutically treat certain genetic diseases. They are set to undergo a clinical trial on a very rare form of blindness called Leber Congenital Amaurosis, where they hope to knockout the gene causing it by injecting CRISPR into the retina.[36] Theoretically, since the disease is so rare, it will be easier to test and get the approved.

Should We?

Amidst this headline grabbing patent battle, in the US alone, the patent office has issued 55 patents regarding CRISPR-related technology to various labs and researchers. The Broad Institute has also received a patent for the CRISPR-Cpf1 system, which is being refined and will be available for public use later in the year.[35] With, this ease-to-edit comes many ethical questions regarding the possible uses of CRISPR. While we are still a long way from using CRISPR as a therapeutic tool in hospitals or to fix genetic disorders before birth, the possibility is being explored, and for some, the hope is that one day it is reached. The main concerns arise with the possibility of germline editing. Somatic editing is when the genes are altered in non-reproductive cells in the body, whereas germline editing is when the egg or sperm cells are altered to create a genetically modified embryo. In 2015, a group of researchers published a paper in Nature that addressed the ethical concerns that need to be considered before jumping into this era of quick and cheap gene editing.[37]

Most of the concern regarding germline edits will have to be addressed in time based on the country and the laws in that particular countries. Different places around the world deal with genome editing differently, with some directly banning any editing to

embryos, and others banning edits unless licensed on a case-to-case basis.[38] When editing a germ-line, alterations are being made to an unborn, consenting human and those changes will affect them for their whole lives, and potentially their offspring's lives as well. The results of edits are ones we do not have the power of fully predicting, because with human embryo research, mosaicism in cell development can dampen or completely alter the outcome of the edit.[37] Another concern is that as CRISPR is explored for therapeutic use, inevitably it will also begin to be exploited for non-therapeutic, commercial use. The 2015 paper in *Nature* stated that germline editing is problematic because, "permitting even unambiguously therapeutic interventions could start us down a path towards non-therapeutic genetic enhancement."[37]

Currently, the laws surrounding genetic research are fairly ambiguous in their language and have not changed fast enough to mirror the advances in science. As new techniques like Cas9 editing emerge, more specific laws and regulations will need to be drawn up to address the mounting ethical concerns that come with editing the human genome. Additionally, issues of the licensing of an edited genome, keeping genetic information, and court cases over unprecedented results could all ensue once we begin to use CRISPR as a viable therapeutic technique. The paper published in 2015 makes a very clear point of the need to open up dialogue on this between the public, the policy-makers and the researchers. They state that "the key to all discussion and future research is making a clear distinction between genome editing in somatic cells and in germ cells."[37] The technique is prepped and clinical trials are underway; it is now simply a matter of time before we have to realistically address whether or not we can safely move forward on the human germline.

[1] Sapranauskas, R.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. The *Streptococcus thermophilus* CRISPR/Cas System Provides Immunity in *Escherichia coli*. *Nucleic Acids Res*. **2011**, *39* (21), 9275–9282.

[2] Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821.

[3] Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339* (6121), 819–823.

[4] Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339* (6121), 823–826.

[5] Baltimore, D.; Berg, P.; Botchan, M.; Carroll, D.; Charo, R. A.; Church, G.; Corn, J. E.; Daley, G. Q.; Doudna, J. A.; Fenner, M.; et al. A Prudent Path Forward for Genomic Engineering and Germline Gene Modification. *Science* **2015**, *348* (6230), 36–38.

[6] Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protocols*. **2013**, *8* (11), 2281–2308.

[7] Miller, J. C.; Holmes, M. C.; Wang, J.; Guschin, D. Y.; Lee, Y.-L.; Rupniewski, I.; Beausejour, C. M.; Waite, A. J.; Wang, N. S.; Kim, K. A.; et al. An Improved Zinc-Finger Nuclease Architecture for Highly Specific Genome Editing. *Nat. Biotechnol*. **2007**, *25* (7), 778–785.

[8] Christian, M.; Cermak, T.; Doyle, E. L.; Schmidt, C.; Zhang, F.; Hummel, A.; Bogdanove, A. J.; Voytas, D. F. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **2010**, *186* (2), 757–761.

[9] Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. V.; Nguyen, N. T.; Zheng, Z.; Gonzales, A. P. W.; Li, Z.; Peterson, R. T.; Yeh, J.-R. J.; et al. Engineered CRISPR-Cas9 Nucleases with Altered PAM Specificities. *Nature* **2015**, *523* (7561), 481–485.

[10] Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Topkar, V. V.; Zheng, Z.; Joung, J. K. Broadening the Targeting Range of *Staphylococcus aureus* CRISPR-Cas9 by Modifying PAM Recognition. *Nat. Biotech*. **2015**, *33* (12), 1293–1298.

[11] Waxmonsky, N. CRISPR 101: Mammalian Expression Systems and Delivery Methods http://blog.addgene.org/crispr-101-mammalian-expression-systems-and-delivery-methods (accessed Apr 2, 2017).

[12] Wang, H.; Russa, M. L.; Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. of Biochem*. **2016**, *85* (1), 227–264.

[13] Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level, 4th Edition*, 4th edition.; Wiley: Hoboken, NJ, 2015.

[14]Trevino, A. E.; Zhang, F. Chapter Eight - Genome Editing Using Cas9 Nickases. In *Methods in Enzymology*; Sontheimer, J. A. D. and E. J., Ed.; The Use of CRISPR/Cas9, ZFNs, and TALENs in Generating Site-Specific Genome Alterations; Academic Press, 2014; Vol. 546, pp 161–174.

[15] Bauer, D. E.; Canver, M. C.; Orkin, S. H. Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9. *J. Visualized Exp.* **2014**, No. 83.

[16] Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; et al. An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nat. Rev. Microbiol.* **2015**, *13* (11), 722–736.

[17] Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163* (3), 759–771.

[18] Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; DeGennaro, E. M.; Winblad, N.; Choudhury, S. R.; Abudayyeh, O. O.; Gootenberg, J. S.; et al. Multiplex Gene Editing by CRISPR-Cpf1 Using a Single crRNA Array. *Nat. Biotech.* **2017**, *35* (1), 31–34.

[19] Yi, L.; Li, J. CRISPR-Cas9 Therapeutics in Cancer: Promising Strategies and Present Challenges. *Biochim. Biophys. Acta, Rev. Cancer* **2016**, *1866* (2), 197–207.

[20] Choi, P. S.; Meyerson, M. Targeted Genomic Rearrangements Using CRISPR/Cas Technology. *Nat. Commun.* **2014**, *5*, 3728.

[21] Maddalo, D.; Manchado, E.; Concepcion, C. P.; Bonetti, C.; Vidigal, J. A.; Han, Y.-C.; Ogrodowski, P.; Crippa, A.; Rekhtman, N.; de Stanchina, E.; et al. In Vivo Engineering of Oncogenic Chromosomal Rearrangements with the CRISPR/Cas9 System. *Nature* **2014**, *516* (7531), 423–427.

[22] Torres, R.; Martin, M. C.; Garcia, A.; Cigudosa, J. C.; Ramirez, J. C.; Rodriguez-Perales, S. Engineering Human Tumour-Associated Chromosomal Translocations with the RNA-Guided CRISPR–Cas9 System. *Nat. Commun.* **2014**, *5*, 3964.

[23] Reardon, S. First CRISPR Clinical Trial Gets Green Light from US Panel. *Nature* **2016**.

[24] Yin, H.; Xue, W.; Chen, S.; Bogorad, R. L.; Benedetti, E.; Grompe, M.; Koteliansky, V.; Sharp, P. A.; Jacks, T.; Anderson, D. G. Genome Editing with Cas9 in Adult Mice Corrects a Disease Mutation and Phenotype. *Nat. Biotechnol.* **2014**, *32* (6), 551–553.

[25] Guan, Y.; Ma, Y.; Li, Q.; Sun, Z.; Ma, L.; Wu, L.; Wang, L.; Zeng, L.; Shao, Y.; Chen, Y.; et al. CRISPR/Cas9-Mediated Somatic Correction of a Novel Coagulator Factor IX Gene Mutation Ameliorates Hemophilia in Mouse. *EMBO Mol. Med.* **2016**, *8* (5), 477–488.

[26] Long, C.; McAnally, J. R.; Shelton, J. M.; Mireault, A. A.; Bassel-Duby, R.; Olson, E. N. Prevention of Muscular Dystrophy in Mice by CRISPR/Cas9-Mediated Editing of Germline DNA. *Science* **2014**, No. 6201, 1184–1188.

[27] Wu, Y.; Liang, D.; Wang, Y.; Bai, M.; Tang, W.; Bao, S.; Yan, Z.; Li, D.; Li, J. Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. *Cell Stem Cell* **2013**, *13* (6), 659–662.

[28] Wu, Y.; Zhou, H.; Fan, X.; Zhang, Y.; Zhang, M.; Wang, Y.; Xie, Z.; Bai, M.; Yin, Q.; Liang, D.; et al. Correction of a Genetic Disease by CRISPR-Cas9-Mediated Gene Editing in Mouse Spermatogonial Stem Cells. *Cell Res.* **2015**, *25* (1), 67–79.

[29] Schwank, G.; Koo, B.-K.; Sasselli, V.; Dekkers, J. F.; Heo, I.; Demircan, T.; Sasaki, N.; Boymans, S.; Cuppen, E.; Van, D. E.; et al. Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. *Cell Stem Cell* **2013**, *13* (6), 653–658.

[30] Li, H. L.; Fujimoto, N.; Sasakawa, N.; Shirai, S.; Ohkame, T.; Sakuma, T.; Tanaka, M.; Amano, N.; Watanabe, A.; Sakurai, H.; et al. Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9. *Stem Cell Rep.* **2015**, *4* (1), 143–154.

[31] Song, B.; Fan, Y.; He, W.; Zhu, D.; Niu, X.; Wang, D.; Ou, Z.; Luo, M.; Sun, X. Improved Hematopoietic Differentiation Efficiency of Gene-Corrected Beta-Thalassemia Induced Pluripotent Stem Cells by CRISPR/Cas9 System. *Stem Cells and Dev.* **2015**, *24* (9), 1053–1065.

[32]Park, C.-Y.; Kim, D. H.; Son, J. S.; Sung, J. J.; Lee, J.; Bae, S.; Kim, J.-H.; Kim, D.-W.; Kim, J.-S. Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. *Cell Stem Cell* **2015**, *17* (2), 213–220.

[33] Liang, P.; Xu, Y.; Zhang, X.; Ding, C.; Huang, R.; Zhang, Z.; Lv, J.; Xie, X.; Chen, Y.; Li, Y.; et al. CRISPR/Cas9-Mediated Gene Editing in Human Tripronuclear Zygotes. *Protein Cell* **2015**, *6* (5), 363–372.

[34] Zou, Q.; Wang, X.; Liu, Y.; Ouyang, Z.; Long, H.; Wei, S.; Xin, J.; Zhao, B.; Lai, S.; Shen, J.; et al. Generation of Gene-Target Dogs Using CRISPR/Cas9 System. *J. Mol. Cell. Biol.* **2015**, *7* (6), 580–583.

[35] McGuire, L. For Journalists: Statement and Background on the CRISPR Patent Process https://www.broadinstitute.org/crispr/journalists-statement-and-background-crispr-patent-process.

[36] Regalado, A. First Human Tests of Gene Editing Could Happen Soon https://www.technologyreview.com/s/543181/crispr-gene-editing-to-be-tested-on-people-by-2017-says-editas

[37] Lanphier, E.; Urnov, F.; Haecker, S. E.; Werner, M.; Smolenski, J. Don't Edit the Human Germ Line. *Nature* **2015**, *519* (7544), 410–411.

[38] Isasi, R.; Kleiderman, E.; Knoppers, B. M. Editing Policy to Fit the Genome? *Science* **2016**, *351* (6271), 337–339.