**CRISPR, a New Way to Modify DNA**

In order to understand what CRISPR is, you would need to know the history of the technique. CRISPR possesses a history that stretches back billions of years but this idea hasn’t really been understood until the end of the 20th century. Once we achieve an understanding of the history, we need to define the mechanics of the process by showing how the pieces work together. Then, we need to look even deeper to see these pieces and how each of them work on a molecular level. Finally, I hope to present significant research in CRISPR from presenting current research discoveries. In this chapter, I present to you what has potential to be the future of genetic modification what has been done with the applications of this technique.

**Introduction: A History Shrouded in Mystery**

First, I want to take you to where this all began with the discovery of this complex system. The setting is the University of Alicante in Alicante, Spain. The year is 1993 and Francisco Mojica is a researcher studying a bacterium called *Haloferax mediterranei* which is known for being an extremophile with a tolerance to high salt concentrations. Mojica is perplexed by a locus of DNA in *H. mediterranei* that are a series of repeated sequences over the course of small pieces of DNA. There is a lot of mystery around these sequences because they are multiple copies that are near perfect matches to one another. These sequences have not been previously classified and do not resemble any other family of repeats. Mojica decided to accept the challenge that these sequences offered and over the span of his career, he would discover that these sequences of DNA are closely related to small sequences found within different types of **bacteriophage**. This was indeed more perplexing to find overlap between a bacterium and their respective virus. An interesting feature that Mojica noted was that, in addition to being perfect copies of one another, these repeated sequences were spaced nearly perfectly as well. Because of this similarity, he was able to hypothesize (accurately) that these sequences would aid in the adaptive immune system. His suspicions would later be confirmed by another group of scientists that discovered this independently around the same time. Clustered regularly-interspaced short palindromic repeats (**CRISPR**) were becoming the talk of the genetic field and every scientist wanted to incorporate CRISPR into their lab work. Being on the forefront of research into these systems, by the year 2000, Mojica was able to find CRISPR systems in 20 different microbes. Some of the organisms that Mojica discovered these loci includes the plague, *Yersinia pestis*, the aerobic pathogen, *Mycobacterium tuberculosis* and the anaerobic bacteria, *Clostridium difficile*.

A key aspect about bacteria is their formidability in their ability to confer resistance. The sheer number of bacteria on this planet is fathomless (estimated to be around a five with thirty zeros after it). With this concept in mind, it would be improbably for there to be a single mechanism that exists that would be able to wipe out an entire species without the bacterium eventually developing a mechanic to resist. If a molecule or compound exists that kills bacteria, they are considered bactericidal. Over the course of billions of years, bacteria were able to confer resistance to some of these **bactericides** secreted by preyed upon organisms, predatory organisms, defense mechanisms by hosts, or naturally hazardous agents. The CRISPR mechanism has garnered a lot of attention be utilized for research needs. The cell has to recognize, react, and respond to foreign nucleic acid and CRISPR is the mechanism to perform this task. CRISPR is a recent scientific advancement that is garnering popularity for its potential to modify the genes of humans, animals, plants, and other organisms with high fidelity. These bacteria would be aided by CRISPR Associated (**Cas**) system that functions to target and cleave viral DNA strands.

Contrary to popular belief, the discovery of the CRISPR systems was bittersweet. Mojica revealed his findings to be rejects numerous times before publication on the basis that they assumed that they fabricated their results. It was so incredulous to believe that there is a mechanism like CRISPR that exists. In 2005, Alexander Bolotin experience a similar process when publishing when studying the recently sequenced *Streptococcus thermophilus* when he discovered that not only did *S. thermophilus* have a CRISPR locus, but it was rather atypical compared to other sequences characterized at the time. This locus lacked some of the previously clarified genes within its locus but instead possessed one thought to configure the activity of a **nuclease**, enzyme that cuts the DNA. This protein was later clarified to be an important CRISPR protein called **Cas9**, which is guided to cleave at a specific site. Bolotin was also able to discover that the spacers between coding regions possessed terminal sequences called the protospacer adjacent motif (**PAM**).

In the timeline of CRISPR discovery, we can see how technology played a large role in describing the system. When CRISPR makes its way to the National Center for Biotechnology Information we meet computational biologist, Eugene Koonin. In 2006, Koonin was able to form computational analysis of the scheme of CRISPR and found that these sequences were associated with the bacterial immune system. These sequences were important and specific in binding to recognizing harmful sequences within viral genomes. This is significant because there were other hypotheses that propose CRISPR to be associate with self-regulation of bacterial cells were need to correct errors in replication. Later in 2008, evidence to support the immune system hypothesis by introducing CRISPR RNAs (**crRNAs**). These crRNAs were hypothesized to function by guiding Cas proteins by binding to viral DNA that is to be targeted for cleavage.

With the knowledge about why the DNA gets cleaved, it is now pertinent to ponder the how. This is exactly what Sylvain Moineau of Laval University in Québec City, Quebec, Canada set out to answer in 2010. The concept that Cas9 targets DNA was supported by Moineau and colleagues, which described that the target of cleavage was site specific. We now know how the target gets determined. It was also found through Moineau’s work that Cas9 was the only protein required for cleavage to occur. We also now know the smoking gun. This are the Type II CRISPR systems which exhibits adaption based upon one large protein (Cas9). Knowing this quality of the system, it was easier to analyze the process of cutting. The research from Moineau’s experience lead to the discovery that the cutting depended on this Cas9 nuclease and that is cut precisely in a blunt-end manner (no overhangs in the double stranded cut). With this major finding, all there is left to do is to apply the research finding.

Thanks to Emmanuelle Charpentier of the Max Planck Society in Berlin, Germany we have a method to the madness. In 2011, Charpentier solidified the mechanism of the CRISPR-Cas9-guided interference. By studying *Streptococcus pyogenes*, a Cas9 oriented bacterium, it was discovered that crRNAs were accompanied by trans-activating CRISPR RNA (**tracrRNA**). These small RNAs formed a complex with crRNA this dimer guides Cas9 to cleave specifically.

**Origin: The Distance of Resistance**

CRISPR is a significant finding that works to aid the bacteria to inactivate foreign viral nucleic acid that impose a threat to the safety of the organism. Cas, a protein that is found universally among CRISPR systems are found in bacteria that are able to protect against foreign viruses, bacteriophages. Bacteria have been around for billions of years, and viruses are assumed to have preluded them. Over time, there was a parasitic relationship where viruses would infect bacterial cells with their RNA for viral protein translation. Well, this was the case until some developed the CRISPR system to retaliate against these foreign invaders.

There are a variety of defense mechanisms for predation as well are invasion within the bacterial genome. The CRISPR system is the bacterial immune defense system against viral invasion. Over time, the bacteria are targeted in masses and in large enough of numbers. The bacteria that would survive would possess a positive, inheritable trait that would persist and be continued to the next generation for resistance to these foreign genetic nucleic acids. Bacteria that would possess these traits would be able to survive in the most competitive of locations by way of their conferred adaptation to viral **lysis**. Bacteria constantly are bombarded with **exogenous** DNA resulting from **conjugation** (transfer of genetic material between two bacterial cells via a bridge**)**, **transduction (**transfer of genetic material between two bacterial cells via a virus**)**, and **transformation (**transfer of genetic material bacteria receiving DNA via genetic predisposition**)**. It would be probable that some bacteria have develop some form of adaptation to unwanted forms of foreign DNA. Hence, we have the CRISPR immune system which is a manner for the bacteria to be able to respond according to future outbreaks of viral infection.

**Structure and Function: The Art of Copy & Paste**

Now, let us delve into some of the most important structures within the CRISPR system. There is a protein associated with CRISPR called CRISPR-associated protein 1, or **cas1**, that is ubiquitously conserved within CRISPR prokaryote immune responses. Usually coupled with the Cas1 is Cas2, which is also found in the various types of CRISPR-Cas system and functions as a spacer between to Cas1 homodimers as a bridge. Cas1 and Cas2, both nucleases, are generally the only proteins conserved in all CRISPR-Cas systems which should emphasize their apparent importance of adaptive immunity via conservation.

The CRISPR locus possesses three major components: these are *cas* genes, a leader sequence, and a repeat-spacer array. *Cas* genes can frequently be associated with CRISPR repeat-spacer arrays and these genes can be separated into over 40 gene families. The locus is structured of repeated sequences, which is short (20-50 base pairs). These repeats typically show dyad symmetry, where they show symptoms of possessing the ability to form a hairpin structure (though they may not be completely palindromic). These repeats are separated by spacers that of a similar length to their corresponding repeat.

There are variations that can be found among the different types of CRISPR systems. In an effort to create minimal CRISPR systems, there has successfully been developed for generating gene knockouts in model organisms (such as bacteria, yeast, fruit flies, mice, and humans). Within *Streptococcus pyogenes*, there is catalytically dead Cas9 (dCas9) that forms a complex with single guide RNAs (**sgRNAs**). sgRNA is a chimeric noncoding type of RNA that possesses three regions: a 20 nucleotide region that is the base pairing sequence, a 42 nucleotide region that forms a hairpin to configure dCas9 binding, and a 40 nucleotide terminator. In order to take advantage of this structure, we simply change the 20 nucleotide region to make a new template for modification. The dCas9 recognizes the hairpin and is directed via this variable region. The next variable condition is the site to which the sgRNA binds which could be simple as using BLAST to confirm the sequence and determine if there may be any unfavorable in the process. For example, as a scientist you don’t want to turn off the gene that causes Alzheimer’s and in the process knockout a structural protein for platelet development. This is a rather simple technique, but this has applications genome-wide in scale. This is of course limited to the availability of dCas9 which is made available by causing two point mutations within the transcript of Cas9 to generate a version of Cas9 that only loses its catalytic function. This is an important adaptation because dCas9 would then be unable to cleave double stranded DNA but still possess the ability to target it. By targeting double stranded DNA specifically, we are able to prevent transcription of a gene that can be found within this region by preventing binding of RNA polymerase.

**Applications: Tying up Loose Genes**

An important milestone in CRISPR history was in 2012, when Siksnys and colleagues purified *S. thermophilus* Cas9-crRNA complex and were able to test the effectiveness of the complex in a test tube. It was found that ***in vitro***, the complex could cleave both strands of DNA targets in the expected PAM regions. These findings were complementary to findings by Moineau and colleagues which were naturally done inside the original cells, ***in vivo***. This was a major feat, however this was eclipsed when they found that they could reprogram the Cas9 cleavage site by using custom designed spacers. These spacers aid the CRISPR system by manipulating crRNAs as well and further showing that they can be about 20 nucleotides in length with no adverse effects on cleavage. This discovery paved the way for genome engineering of RNA-guided DNA endonucleases. Eventually, Charpentier with the assistance of her new colleague, Jennifer Doudna, would clarify the controversy that could arise by manipulating human genomes. Collaborating together, Charpentier and Doudna were able to use Cas9, crRNA, and tracrRNA to show evidence that Cas9 could cut pure DNA *in vitro*. Not only could they show this, but they could also show that the manipulation of the crRNA to custom-designed variant controlled the specificity of the cut. And finally, they showed that crRNA and tracrRNA could fuse to become sgRNA that would grow to become widely used in genome editing with modifications tailored to the experiment.

In 2012, a researcher by the name of Feng Zhang was able to target 16 different sites within the human and mouse genomes to directly cause modifications with high efficiency and accuracy. Zhang found that by using non-homologous end-joining, deletions could occur and with homologous recombination, insertions are possible. These genome editing systems were able to adapt CRISPR-Cas9 system. With modification, this can be further programmed to target longer or multiple genome loci. CRISPR/Cas9 can occur by using plasmid to transfect a cell that needs modification. The plasmid gets expressed and encodes for Cas9, tracrRNAs, and crRNAs, all important components for genome editing via CRISPR/Cas9. There are a multitude of potential applications available for the application of CRISPR. Of those, the most impactful use would be the ability to inactivate potentially harmful genes in humans.

In recent developments of CRISPR, it is an idea that has garnered the attention of much of the scientific population. With the popularity, researchers were eager to present their research that displayed that CRISPR gene editing could range among many organisms (yeast, nematodes, fruit flies, zebrafishes, mice, and monkeys). As talk about the possibilities within humans were introduced, so were the argument for dissent. There were social concerns, with there being debates about the ethical implications of designer babies. For example, there is difference in choosing a baby that does not possess the parent’s inheritable disease because prior to conception is was inactivated versus a baby that was chosen based on aesthetics such as hair/eye color, height, or possibly intelligence. An important question that should arise of course is the reason to why CRISPR is being experimented and why we should move forward in applying the wealth of knowledge displayed before us. Though some may accuse scientists to be on a journey to edit the human genome or to even study human disease, this was not the case. Scientists were unsure of what they found originally and they were driven by the curiosity factor. There wasn’t room for hypothesis in the discovery of CRISPR in terms of potential applications, but it was certainly a result. As large as a discovery CRISPR is to the science of gene editing, the central role this played was to show how important it is for scientists to be honest to the science even when they may not be published for it. Also, CRISPR enlightens that the mystery is only as great as the questions that are proposed and finally that even a seemingly underwhelming series of scientific advancements could lead to the central game changing discovery.

**Glossary**

**bactericide** – a substance that kills bacteria.

**bacteriophage** – a virus that specifically targets a bacterial cell for infection.

**Cas** – (**C**RISPR **as**sociated protein). An enzyme that cleaves DNA that are associated with CRISPR that primarily is used as a defense mechanism against foreign bacteriophages.

**Cas9** – (**C**RISPR **as**sociated protein **9**). An enzyme that is guided by RNA to function as a DNA endonuclease for CRISPR immune response.

**conjugation –** The transfer of DNA, in the form of a plasmid, between a singular bacterial cell to another bacterial cell.

**CRISPR** – (**C**lustered **R**egularly-**I**nterspaced **S**hort **P**alindromic **R**epeats). Gene editing that is guided by RNA that uses Cas to modify specific regions within the genome.

**crRNAs** – Small pieces of RNA found within CRISPR utilizing microbes, aid Cas protein by binding to DNA.

**exogenous** – Originating from outside of the organism. (*ant.* endogenous)

***in vivo*** – Of or occurring inside of the cell. (*ant. in vitro*)

***in vitro*** – Of or occurring outside of the cell. (*ant. in vivo*)

**leader sequence** – a region of an mRNA, or DNA that encodes for such, that is directly upstream from the initiation codon. In CRISPR, this sequence is upstream of the PAM.

**lysis** – Rupturing of a cell that occurs after membrane weakening (typically by enzymes, viruses, or osmosis)

**nuclease** – An enzyme that functions by cleaving DNA or RNA via its phosphodiester bonds between nucleic acids.

**PAM** – (**P**rotospacer **A**djacent **M**otifs). A sequence of DNA that is found immediately after DNA sequences that is recognized by *Cas9* endonuclease.

**sgRNA** – (**s**ingle **g**uide **R**NA) RNA that is targeted by a Cas9 protein to guide Cas9 to perform snipping of DNA.

**tracrRNA** – (**tr**ans-**a**ctivating **CR**ISPR **R**NA). Small pieces of RNA found within CRISPR utilizing microbes that bind to *crRNAs* to aid Cas9 binding specificity.

**transduction –** The process of transferring DNA to a cell via a virus.

**transformation –** The uptake of foreign DNA by means of genetic manipulation of the cell (*ex.* Electroporation).

**References**

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). Molecular biology of the cell. Chicago, Illinois. Garland Science.

2. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315(5819), 1709-1712.

3. DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336–4343.

4. Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J. and O’Connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, *194*(4), pp.1029-1035.

5. Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science, 327(5962), 167-170.

6. 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), 816-821.

7. Lander, E. S. (2016). The Heroes of CRISPR. Cell, 164(1), 18-28.

8. Ledford, H. (2015). CRISPR, the disruptor. Nature, 522(7554), 20-24.

9. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), pp.823-826.

10. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). CRISPR-assisted editing of bacterial genomes. *Nature Biotechnology*, *31*(3), 233–239.

11. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), pp.935-949.

12. Nuñez, 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), 528-534.

13. 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(3), 653-663.

14. Swarts, D. C., Mosterd, C., Van Passel, M. W., & Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. PloS one, 7(4), e35888.

15. Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, *153*(4), 910-918.