

The Journey of CRISPR-Cas

“Playing God with Life”

Dr. Devashish Rath*

The genetic material, which happens to be DNA in vast majority of organisms, is the molecular blue-print of life. The genetic code that resides as a sequence in the DNA is first copied in the form of RNA and then further translated as proteins. The proteins carry out the structural or biochemical functions in a cell. In 1953, J.D. Watson and F.H.C. Crick reported the molecular structure of DNA [1]. Ever since, scientists have tried to develop technologies that can manipulate the genetic material of cells and organisms. The genetic manipulations become increasingly complex and difficult to achieve as we move from lower organisms such as bacteria to higher organisms such as humans. Many organisms have proved genetically intractable as genetic manipulation in these remained elusive. With the discovery of the RNA-guided CRISPR-Cas9 system, an easy and effective method for genome engineering has now become a reality. The development of this technology has enabled scientists to modify DNA sequences in a wide range of cells and organisms making it possible to change the code of life. Genomic manipulations are no longer an experimental bottleneck. Today, CRISPR-Cas9 technology is used widely in basic science, biotechnology and in the development of future therapeutics [2]. French microbiologist Emmanuelle Charpentier, the Director of the Max Planck Unit for the Science of Pathogens, Berlin, Germany, and the American biochemist Jennifer A. Doudna, Professor at the University of California, Berkeley, USA and Investigator, Howard Hughes Medical Institute shared the Nobel Prize in Chemistry, 2020 “for the development of a method for genome editing. The tool for genome editing came from the study of CRISPR-Cas9 system from a human pathogen named *Streptococcus pyogenes*.

In the ensuing text I will recount the historical development of this exciting field of science, appropriately juxtaposing the contributions of the aforementioned Nobel laureates which culminated in the development of an elegant yet powerful tool for genome engineering.

The discovery of the CRISPR-Cas system in prokaryotes

Starting with the bacterium *Escherichia coli* (*E. coli*) in 1987 and in subsequent years unusual repeated elements were observed in the genomes of some bacteria. As more and more genome sequences accumulated in the databases, bioinformatics analyses revealed that such structures were common in the genomes of bacteria and these were characterized by common peculiar features: clusters of a short (about 25-50 bp), partially palindromic elements (repeats) separated by unique intervening sequences of constant length (spacers) [3-5]. Based on these features the acronym CRISPR (clustered regularly interspaced short palindromic repeats) was coined for these elements [6]. Bioinformatic analyses subsequently identified a set of genes occurring exclusively in CRISPR containing bacteria and always located adjacent to the CRISPRs aptly named as cas (CRISPR-associated) genes [6]. These genes mostly encoded proteins associated with DNA metabolism. It was also observed that CRISPR locus (repeats and spacers) gave rise to RNA called as CRISPR RNA (crRNA) but no proteins. The biological function of the CRISPR and cas genes remained a mystery until 2005, when three groups of researchers working independently reported that the unique intervening parts (spacers) of the CRISPRs were derived from bacteriophages (viruses that attack bacteria) and plasmids [7-9]. This suggested that CRISPR-cas are probably involved in protection from invading DNA such as viruses and plasmids in bacteria.



Born 1968 in Juvisy-sur-Orge, France. Ph.D. 1995 from Institut Pasteur, Paris, France. Director of the Max Planck Unit for the Science of Pathogens, Berlin, Germany.

Emmanuelle Charpentier



Born 1964 in Washington, D.C, USA. Ph.D. 1989 from Harvard Medical School, Boston, USA. Professor at the University of California, Berkeley, USA and Investigator, Howard Hughes Medical Institute.

Jennifer A. Doudna

Nobel Prize Winners of 2020 (Chemistry)
“for the development of a method for genome editing”

Photo credit @ <https://www.whatisbiotechnology.org/index.php/people/summary/Doudna>
<https://www.forbes.com/profile/emmanuelle-charpentier/?sh=3853b4027da2>

CRISPR-Cas functions as an adaptable defence system

The notion that CRISPR-Cas could work as anti-virus defence was confirmed in 2007 [10] from a collaborative effort of scientists working with Danisco, a company making dairy products and scientists in academia. The Yoghurt making bacterium *Streptococcus thermophilus* (containing a class 2 CRISPR type) was infected with a virus. Resistant bacteria that survived the virus attack were isolated and their CRISPR locus was studied. The resistant bacteria had acquired new spacer sequences, which matched DNA sequence within the infecting virus. Deletion of the spacer sequence led to loss of

*The author is Group Leader, CRISPR Biology Group Applied Genomics Section, BARC

CRISPR and cas genes

The biological function of the CRISPR and cas genes remained a mystery until 2005, when three groups of researchers working independently reported that the unique intervening parts (spacers) of the CRISPRs were derived from bacteriophages (viruses that attack bacteria) and plasmids.

Genomic manipulations are no longer an experimental bottleneck.

Today, CRISPR-Cas9 technology is used widely in basic science, biotechnology and in the development of future therapeutics.

resistance and conversely insertion of spacer sequence in the genome of sensitive bacteria made them resistant to the virus. Furthermore, inactivation of one of the cas genes (cas9) resulted in loss of virus resistance. These experiments established two things, a) the specificity of resistance to a particular virus was dependent on the spacer sequence of the CRISPR locus and b) a gene cas9 was involved [10]. However, the molecular details of the working of the CRISPR system was still missing.

Glimpses into how the CRISPR-Cas system works came from investigations of *E. coli*, which contains a Class 1 CRISPR-Cas system comprising of eight different cas genes. Researchers led by Prof. John van der Oost at Wageningen University, Netherlands established that proteins made from five of these genes formed a multiprotein complex termed Cascade (CRISPR-associated complex for antiviral defence). Cascade was shown to cut the RNA made from the CRISPR locus inside the repeat sequence to produce shorter crRNA molecules containing the virus-derived sequence [11]. These mature crRNA molecules (having sequence complementary to viral DNA sequence) served as guide molecules that enabled Cascade assisted by another protein Cas3 to identify the infecting virus DNA and destroy it. These results suggested that CRISPR function operated in two steps: first, making of Cas proteins and processing of crRNA to mature form and formation of a complex between them, and second, an interference step in which the viral DNA was recognized and destroyed. These experiments also showed that by providing crRNA matching the viral DNA and Cas proteins a bacterium could be protected from virus attack [11].

By the year 2011, based on the work described above and work from other labs, it was realized that a number of different CRISPR-Cas systems existed in bacteria, which are now divided into two major classes [12]. In the class 1 systems, multiple Cas proteins assemble into a large CRISPR-associated complex for antiviral defence (Cascade). In contrast, class 2 systems are simpler and contain a single crRNA-binding protein (e.g. Cas9) that has all the functions necessary for viral interference. Studies had also established that the CRISPR-Cas system functioned at three different steps: (i) integration of new spacer DNA sequences into CRISPR loci (adaptation phase), (ii) processing and formation of mature crRNAs (expression phase),

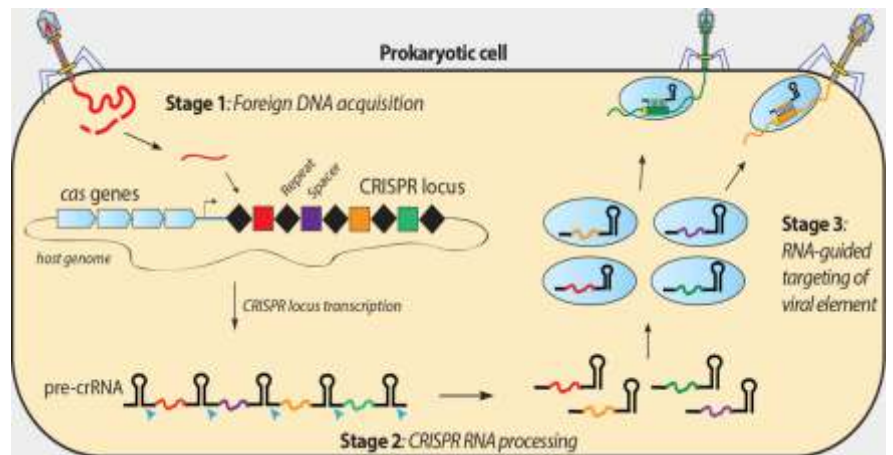


Fig. 1: A general scheme for the function of the CRISPR-Cas adaptive immune system.

Three stages are identified. Stage 1 (Adaptation): Short fragments of double-stranded DNA from a virus or plasmid are incorporated into the CRISPR array on host DNA. Stage 2 (crRNA Maturation): Pre-crRNA are produced by transcription and then further processed into smaller crRNAs, each containing a single spacer and a partial repeat. Stage 3 (Interference): When crRNA recognize and specifically base-pair with a region on incoming plasmid or virus DNA the DNA is cut by associated Cas nuclease. Interference can be separated both mechanically and temporally from CRISPR acquisition and expression.

Photo credit @ https://doudnalab.org/research_areas/crispr-systems/

and (3) recognition and destruction of foreign DNA (interference phase) [13,14] (Fig. 1).

Discovery of tracrRNA and its role in crRNA maturation

In 2011, Emmanuelle Charpentier and colleagues working on an infectious bacterium *Streptococcus pyogenes* [15] identified a class 2 CRISPR locus with an adjacent cas9 gene. Apart from pre-crRNA and mature crRNA they identified a small RNA species (tracrRNA) that contained a stretch of 25 nucleotides (nt) with almost perfect complementarity to the repeat regions of the CRISPR locus. Charpentier hypothesized that tracrRNA formed a duplex with pre-crRNA [15] and the structure of RNA duplex could act as processing sites for an enzyme endoribonuclease (enzyme that cuts RNA) RNase III present in the bacterium. Basically Charpentier proposed that the two RNAs are co-processed upon pairing and neither would be processed alone. The idea was proved by showing that deletion of the tracrRNA prevented pre-crRNA processing and vice versa. Further they went on to demonstrate that RNaseIII could process a heteroduplex formed between tracrRNA and pre-crRNA in vitro and was required for tracrRNA and pre-crRNA processing in vivo. Finally, they also proved that the processing required Cas9 protein. This work identified all the elements involved in crRNA maturation in a simpler

“By simply introducing an engineered sgRNA and the Cas9 nuclease, scientists are now able to make precise changes in the genome.

class 2 system and hypothesized that once processed with the help of tracrRNA the mature crRNA was sufficient to guide the cleavage of the target DNA by the Cas9 nuclease.

Towards a Genome Editing Tool

Emanuelle Charpentier collaborated with Jennifer A. Doudna to investigate if crRNA could be used to program the sequence specificity of Cas9. In other words, the idea was to provide a crRNA, that is complementary to a pre-determined target DNA, to make Cas9 recognize and cut that target in vitro. To their surprise, addition of crRNA to purified Cas9 did not result in target DNA cleavage [15,16]. After some failed experiments, in a 'eureka' moment the researchers added the tracer RNA to the in vitro reaction and it triggered the cutting of the target DNA by Cas9. Thus, the tracrRNA not only was required for pre-crRNA processing by the enzyme RNase III but was also critical for activating crRNA-guided DNA cleavage by Cas9. Through a series of biochemical experiments, the researchers showed that Cas9 had two domains and each were shown to cleave one strand of the DNA producing a double stranded break in the DNA. Further, they delineated the regions of tracrRNA and crRNA that were absolutely essential for Cas9-catalyzed cutting of the DNA through serial deletions. Based on these results the authors proceeded to simplify the system by capturing the structural features of tracrRNA and crRNA in a single molecule. Indeed, they could fuse them into a chimeric single-guide RNA (sgRNA) molecule that could act along with Cas9 (Fig.2). Finally, they also demonstrated that sgRNA sequence can be changed to make Cas9 target DNA of interest. So, they had created a simple two-component system that could be programmed by changing the sgRNA sequence to target and cut any DNA of choice (Fig.2). The importance of this finding was highlighted in the abstract of the paper where the authors wrote: “Our study reveals a family of endonucleases that use dual-RNAs for site-specific DNA cleavage and highlights the potential to exploit the system for RNA programmable genome editing” [16].

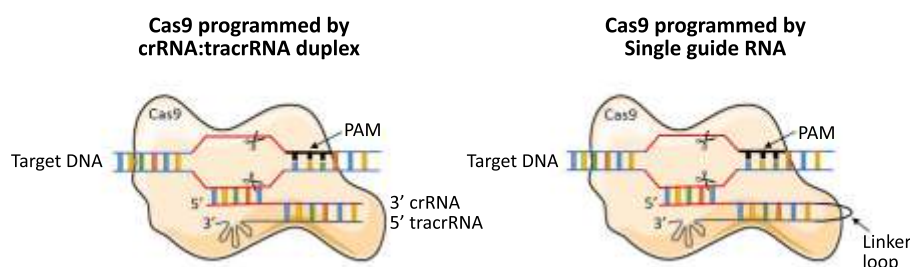


Fig. 2: Cutting of target DNA by Cas9 requires two small RNAs crRNA (green) and tracrRNA (red). These two RNAs can be combined into a single RNA called guide RNA (gRNA) which captures structural features of both.

A Lithuanian biochemist Virginijus Siksnys and his colleagues reached similar conclusions working with the *Streptococcus thermophilus* CRISPR-Cas9 system and in fact communicated them to the journal Cell before Charpentier and Doudna, however, their paper was not accepted. Subsequently, they wrote to the PNAS, USA where it underwent the usual review process before acceptance. While Siksnys and colleagues lost crucial time in the process, Charpentier and Doudna communicated their findings to the journal Science which accelerated the process of the review and publication thereby enabling them to publish their findings ahead of Siksnys. As in Charpentier and Doudna's work, Siksnys and colleagues also demonstrated that Cas9 cleaves the target DNA, that cleavage specificity is directed by the crRNA sequence, and that the two nuclease domains within Cas9, each cleave one strand. However, the researchers did not notice the crucial importance of tracrRNA for sequence-specific cleavage of target DNA [17].

The impact and application of CRISPR-Cas genome editing technology

Though the in vitro experiments of Charpentier and Doudna, published in 2012, had established the potential of CRISPR-Cas9 for genome editing, the in vivo experimental demonstration was lacking. In the year 2013, two groups achieved the stupendous feat of genome editing in human and mouse cells with CRISPR-Cas9 and independently published the results in the same issue of the journal Science [18,19]. These landmark studies demonstrated that Cas9 nucleases could be directed by crRNA of a defined sequence to induce precise cuts in the genes of mouse and human cells. Since then the application of CRISPR-based genome editing in different organisms has exploded. By simply introducing an engineered sgRNA and the Cas9 nuclease, scientists are now able to make precise changes in the genome (Fig. 3). This has allowed researchers to understand what functions genes do, find out the changes in genes that are associated with disease, introduce new functions into bacteria for biotechnological or industrial applications.

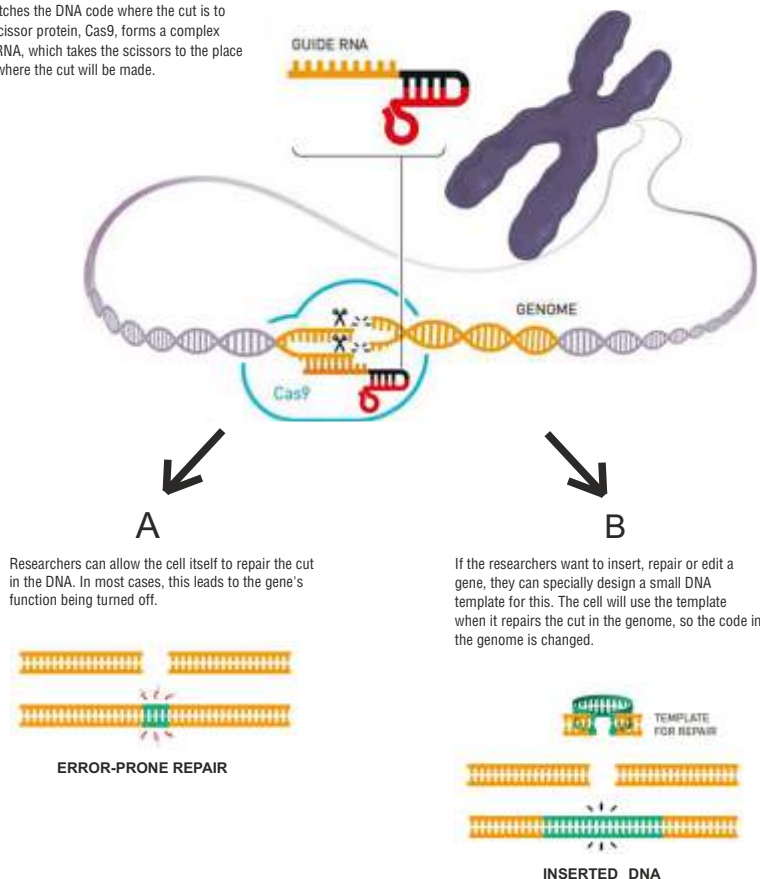
Some of the areas where CRISPR-mediated genome editing technology is beginning to have a tremendous impact is improvement of crops and treatment of genetic diseases like sickle cell disease and b-thalassemia. Recently it has been shown that CRISPR technology can be used to edit gene in human embryos which has raised grave concerns on ethical and social issues. There is a growing recognition that the technology needs to be regulated for responsible use so that social, legal, ethical and scientific challenges arising out of human genome editing can be adequately addressed.

Technology regulation

There is a growing recognition that the technology needs to be regulated for responsible use so that social, legal, ethical and scientific challenges arising out of human genome editing can be adequately addressed.

The CRISPR/Cas9 genetic scissors

When researchers are going to edit a genome using the genetic scissors, they artificially construct a guide RNA, which matches the DNA code where the cut is to be made. The scissor protein, Cas9, forms a complex with the guide RNA, which takes the scissors to the place in the genome where the cut will be made.



@ Johan Jarnestad/The Royal Swedish Academy of Sciences

Fig. 3: Genome editing with Cas9

The Cas9 enzyme is directed to target DNA by a guide RNA and produces a double-stranded break. The break can be repaired by cellular machinery in an error-prone manner leading to gene knock-out of the function of the gene (A). Or a piece of DNA can be inserted during repair leading to disruption of the gene or replacement of a part (B).

Photo credit @ The Royal Swedish Academy of Science

Reminiscences



Devashish Rath

As I was working on the class 1 CRISPR-Cas system in *Escherichia coli* during my post-doc tenure at the Uppsala University, Sweden, I had the privilege of interacting with both the Nobel laureates. I visited the laboratory of Dr.

Charpentier in Umea University, Sweden in the year 2012 just after the publication of their tracrRNA work. Dr. Charpentier had returned after attending a 'CRISPR conference' in USA and so I had many questions regarding the work presented in the conference. Dr. Charpentier jocularly remarked that "My Swedish friend is very inquisitive". I reminded her that I belonged to India and was in Sweden for a transitory period. It is heart-warming to note that I received a warm welcome in Charpentier lab and had a lively discussion with her and other members of her group. The very next year I met Dr. Doudna, in a 'CRISPR conference' at St. Andrews, Scotland. I was presenting my work on developing a gene silencing application using a type I CRISPR system which immediately attracted the interest of Dr. Doudna as she had recently worked on a similar application using the Cas9 system. In the ensuing discussion she gave me many useful suggestions on enhancing the efficiency of the system. I must acknowledge that I owe these special moments to my then post-doctoral mentor Dr. Magnus Lundgren. While working with the team of Prof. John van der Oost, Dr. Lundgren contributed to seminal work in *E. coli* that provided first molecular insights into the working of the CRISPR-Cas system for which he was a well-recognized name in the CRISPR community.

Acknowledgements:

The author wishes to thank Shri Mandeep Singh, SO/C, AGS, BSG for help with compilation of figures. The author is grateful to Dr. S. K. Ghosh, Ex-Associate Director, BSG, BARC and Dr. G. Ravi Kumar, Head, SIRD and Dr. A.V.S.S.N. Rao, Head, AGS, BSG for their encouragement. The author thanks Dr. Santosh Kumar S., SO/H, RB&HSD for his suggestions to improve the article.

Corresponding Author*

Dr. Devashish Rath (devrath@barc.gov.in)

References

- [1] Watson, J.D. and Crick, F.H. "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid". *Nature*, **171(4356)**, (1953): 737-38.
- [2] Knott, G.J. and Doudna, J.A. "CRISPR-Cas guides the future of genetic engineering". *Science*, **361(6405)**, (2018): 866-69.
- [3] Ishino, Y., et al., "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product". *J Bacteriol*, **169(12)**, (1987): 5429-33.
- [4] Mojica, F.J., Juez, G. and Rodriguez-Valera, F. "Transcription at different salinities of *Haloflex mediterranei* sequences adjacent to partially modified PstI sites". *Mol Microbiol*, **9(3)**, (1993): 613-21.
- [5] Mojica, F.J., et al., "Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria". *Mol Microbiol*, **36(1)**, (2000): 244-46.
- [6] Jansen, R., et al., "Identification of genes that are associated with DNA repeats in Prokaryotes". *Mol Microbiol*, **43(6)**, (2002): 1565-75.
- [7] Mojica, F.J., et al., "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements". *J Mol Evol*, **60(2)**, (2005): 174-82.
- [8] 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 (Reading)*, **151(Pt 3)**, (2005): 653-63.
- [9] Bolotin, A., et al., "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin". *Microbiology (Reading)*, **151(Pt 8)**, (2005): 2551-61.
- [10] Barrangou, R., et al., "CRISPR provides acquired resistance against viruses in prokaryotes". *Science*, **315(5819)**, (2007): 1709-12.
- [11] Brouns, S.J., et al., "Small CRISPR RNAs guide antiviral defense in prokaryotes". *Science*, **321(5891)**, (2008): 960-64.
- [12] Makarova, K.S., et al., "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants". *Nat Rev Microbiol*, **18(2)**, (2020): 67-83.
- [13] Bhaya, D., M. Davison, and R. Barrangou, "CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation". *Annu Rev Genet*, **45**, (2011): 273-97.
- [14] Terns, M.P. and Terns, R.M. "CRISPR-based adaptive immune systems". *Curr Opin Microbiol*, **14(3)**, (2011): 321-27.
- [15] Deltcheva, E., et al., "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III". *Nature*, **471(7340)**, (2011): 602-7.
- [16] Jinek, M., et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity". *Science*, **337(6096)**, (2012): 816-21.
- [17] Gasiunas, G., et al., "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria". *Proc Natl Acad Sci U S A*, **109(39)**, (2012): E2579-86.
- [18] Cong, L., et al., "Multiplex genome engineering using CRISPR/Cas systems". *Science*, **339(6121)**, (2013): 819-23.