



Scientific Background on the Nobel Prize in Chemistry 2020

## A TOOL FOR GENOME EDITING



## A Tool for Genome Editing

*The Royal Swedish Academy of Sciences has decided to award **Emmanuelle Charpentier** and **Jennifer A. Doudna** the Nobel Prize in Chemistry 2020, for the development of a method for genome editing.*

### Introduction

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. 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. 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].

### The discovery of the CRISPR-Cas system in prokaryotes.

The work that eventually led to the discovery of the powerful CRISPR-Cas9 system for genome editing began with the identification of repeated genome structures present in bacteria and Archaea. In 1987, a report noted an unusual repeated structure in the *Escherichia coli* genome, which contained five highly homologous sequences of 29 base pairs (bp), including a dyad symmetry of 14 bp that were interspersed by variable spacer sequences of 32 bp [3]. Some years later, similar, repeated structures were identified in the genome of the halophilic Archaea *Haloferax mediterranei*, with 14 almost perfectly conserved sequences of 30 bp, repeated at regular distances [4].

Subsequent bioinformatics analyses revealed that these types of repeats were common in prokaryotes and all contained the same peculiar features: a short, partially palindromic element occurring in clusters and separated by unique intervening sequences of constant length, suggesting an ancestral origin and high biological relevance [5]. The term CRISPR was introduced, an abbreviation for clustered regularly interspaced short palindromic repeats [6].

An important step towards understanding the function of CRISPR came with the identification of CRISPR-associated (*cas*) genes, a group of genes only present in CRISPR-containing prokaryotes and always located adjacent to CRISPRs. The identified *cas* genes encoded proteins with helicase and nuclease motifs suggesting a role in DNA metabolism or gene expression [6]. The association with CRISPR was used as a defining characteristic and over the coming years a number of Cas protein subfamilies were described [7, 8].

The functional importance of the CRISPR loci remained elusive until 2005, when researchers noted that the unique CRISPR sequences were derived from transmissible genetic elements, such as bacteriophages and plasmids [9-11]. Prokaryotes carrying these specific sequences appeared protected from infection, since plasmids or viruses containing a sequence matching a spacer (named protospacers) were usually absent in the prokaryote carrying the spacer [9, 11].

These correlative findings suggested a function for CRISPRs in prokaryotic defence against invading foreign DNA and the spacer sequences were described as a ‘memory of past “genetic aggressions”’ [10]. It had already been shown that CRISPRs were transcribed into long RNA



molecules (pre-crRNA), which were subsequently processed by cleavage within the repeat sequences to yield small CRISPR-RNAs (crRNAs) [4, 12]. Taken together these observations indicated that crRNA could play a role in targeting viral nucleic acids, perhaps in a manner similar to RNAi in eukaryotic cells. It was also hypothesized that the Cas proteins was involved in this process [9].

Later research has indeed demonstrated that crRNA binds to one or more Cas proteins to form an effector complex that targets invading nucleic acids. Extensive efforts during the past 25 years have identified a number of different CRISPR-Cas systems, which are now divided into two major classes [13]. In the Class 1 systems, specialised Cas proteins assemble into a large CRISPR-associated complex for antiviral defence (Cascade). The Class 2 systems are simpler and contain a single multidomain crRNA-binding protein (e.g. Cas9) that contains all the activities necessary for interference.

### **CRISPR-Cas functions as an adaptable defence system**

The hypothesis that CRISPR-Cas systems could confer resistance to invading foreign DNA was verified in 2007 [14]. In an elegant set of experiments, scientists studied a Class 2 system in a strain of *Streptococcus thermophilus*, which they infected with virulent bacteriophages. Next, bacteria resistant to infection were isolated and their CRISPR loci analysed. The experiment revealed that resistant bacteria had acquired new spacer sequences, which matched sequences within the infecting phage used to select resistance. Deletion of the spacer region led to loss of resistance, and the phages that were able to grow on resistant bacteria had accumulated mutations in the protospacer sequence in the phage genome. Furthermore, inactivation of one of the *cas* genes (*cas5*) resulted in loss of phage resistance. The experiments thus demonstrated a role for *cas* gene products in CRISPR-Cas-mediated immunity and that the specificity of the system was dependent on the spacer sequences [14].

Further insights into the function of CRISPR-Cas came from investigations of *E. coli*, which contains a Class 1 CRISPR-Cas system encoding no less than eight different Cas proteins. Five of these gene products could be purified as a multiprotein complex termed Cascade (CRISPR-associated complex for antiviral defence). Cascade was shown to function in pre-crRNA processing, cleaving the long transcripts in the repeated regions and thereby producing shorter crRNA molecules containing the virus-derived sequence [15]. After cleavage, the mature crRNA molecules were retained by Cascade, and, assisted by a *cas*-encoded helicase, Cas3, they served as guide molecules that enabled Cascade to interfere with phage proliferation. The results thus suggested two different steps in CRISPR function: first, CRISPR expression and crRNA maturation, and second, an interference step that required the Cas3 protein. The results also provided evidence suggesting that the *E. coli* CRISPR-Cas system targets phage DNA and not RNA, inasmuch as crRNA with complementarity to either of the two DNA strands could interfere with phage proliferation [15].

Conclusive evidence for DNA being the target of CRISPR-Cas interference came from elegant experiments using a strain of *Staphylococci epidermidis* that contained a CRISPR array with a spacer sequence homologous to a gene present in a conjugative plasmid [16]. Transfer of the plasmid into the strain occurred only if the spacer sequence was mutated or deleted. A self-splicing intron was inserted into the target sequence on the plasmid. In this way, the CRISPR spacer would be complementary not to the DNA, as it is disrupted by an intron, but to the RNA, which would be spliced, reconstituting the sensitive target. Indeed, insertion of the self-splicing intron was sufficient to overcome CRISPR-Cas inhibition of plasmid transfer, strongly implicating



DNA as the primary target [16]. This conclusion was further supported from studies of *S. thermophilus*, in which the CRISPR-Cas system was shown to cleave both bacteriophage and plasmid DNA *in vivo* [17].

### **Protospacer adjacent motifs distinguish CRISPR from invading DNA.**

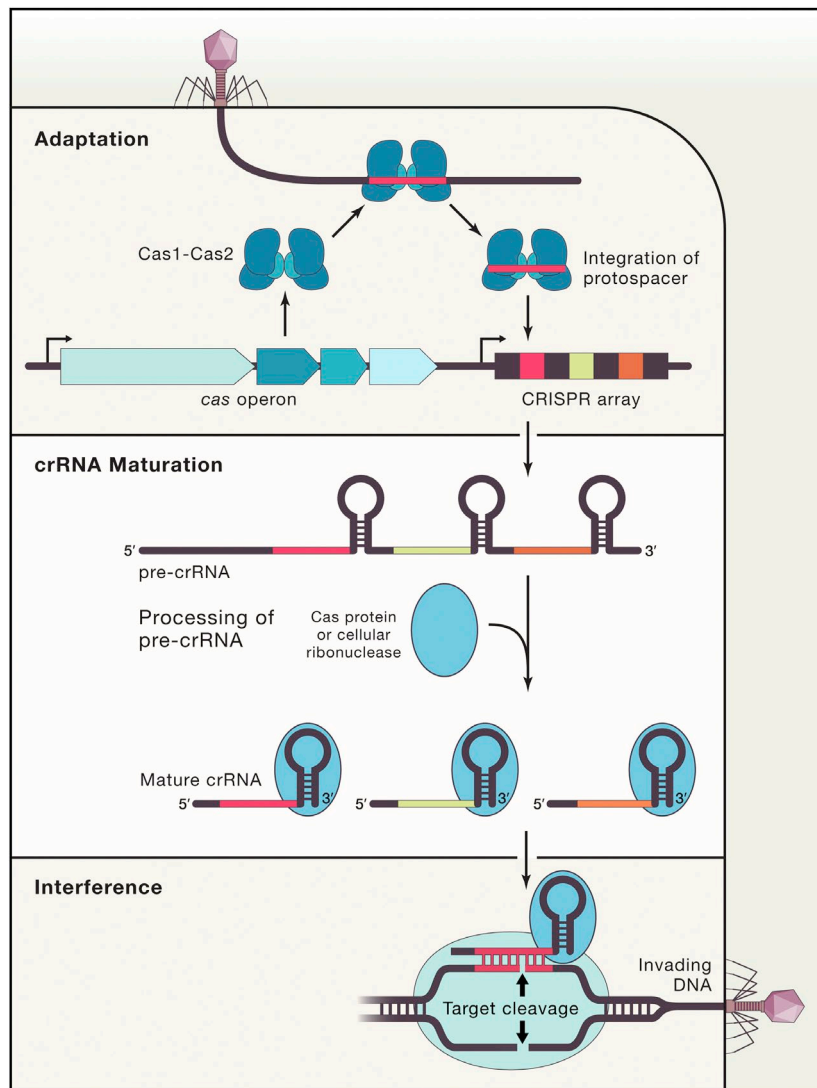
If spacers lead to cleavage of DNA with matching sequences, how do they avoid cleaving their own CRISPR spacers? The answer to this question came from studies of sequences around protospacers, i.e. the sequences in the phage genomes that had given rise to spacers. Short sequence motifs were noted just a couple of nucleotides away from protospacer sequences [11, 18]. These motifs were later labelled protospacer adjacent motifs or PAMs [19].

The functional importance of PAMs became clear from work studying the phage response to CRISPR-encoded resistance in *S. thermophilus*. In these studies, phages that had overcome bacterial resistance were isolated and analysed. These studies revealed that a number of those resistant to CRISPR immunity had acquired mutations in the PAMs, implicating these short sequences as important for targeting [20]. Later studies have demonstrated that the PAM sequences are required both for target interference and for uptake of new spacer sequences into CRISPRs [21, 22].

### **Discovery of the CRISPR-Cas9 system**

By 2011, it was clear that CRISPR-Cas systems were widespread in prokaryotes and functioned as adaptive immune systems to combat invading bacteriophages and plasmids (Figure 1). Studies had also established that the Cas proteins functioned at three different levels: (i) integration of new spacer DNA sequences into CRISPR loci, (ii) biogenesis of crRNAs, and (3) silencing of the invading nucleic acid [23, 24].

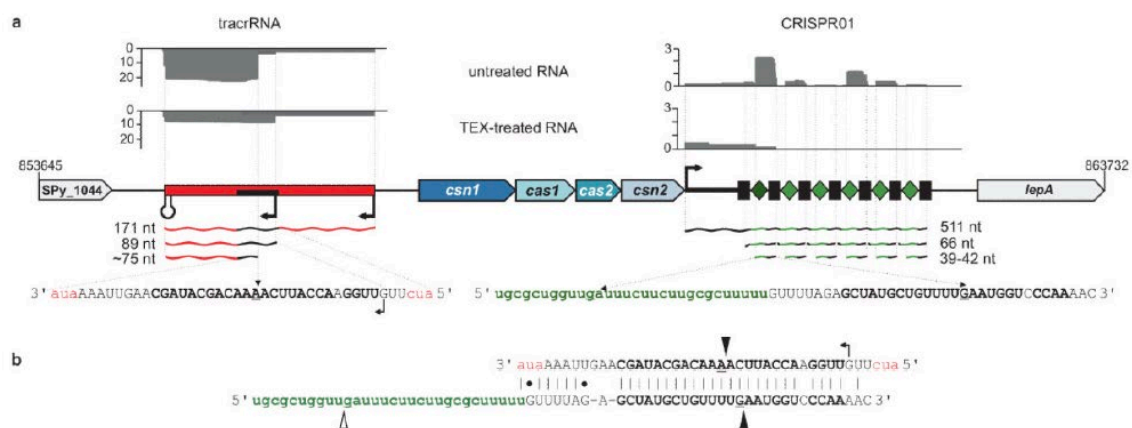
The identification of CRISPR-Cas9 as a tool for genomic editing came from studies of the Class-2, Type-II CRISPR-Cas system in *S. thermophilus* and the related human pathogen *Streptococcus pyogenes*. This system contains four *cas* genes, three of which (*cas1*, *cas2*, *csn2*) are involved in spacer acquisition, whereas the fourth, *cas9* (formerly named *cas5* and *csn1*), is needed for interference [14]. In support of this notion, inactivation of the *cas9* gene prevented cleavage of target DNA [17]. To further define the elements required for immunity, the *S. thermophilus* CRISPR-Cas system was introduced into *E. coli*, where it provided heterologous protection against infection with phages and plasmids [25]. Using this experimental model, parts of the system were inactivated to define the components required for protection. The work clearly demonstrated that the Cas9 protein alone was sufficient for the CRISPR-encoded interference step, and that two nuclease domains present in the protein, HNH and RuvC, were both required for this effect [25].



**Figure 1.** A general scheme for the function of the CRISPR-Cas adaptive immune system as presented in [26]. Three stages are identified. Adaptation: Short fragments of double-stranded DNA from a virus or plasmid are incorporated into the CRISPR array on host DNA. crRNA Maturation: Pre-crRNA are produced by transcription and then further processed into smaller crRNAs, each containing a single spacer and a partial repeat. Interference: Cleavage is initiated when crRNA recognize and specifically base-pair with a region on incoming plasmid or virus DNA. Interference can be separated both mechanistically and temporally from CRISPR acquisition and expression.

## Discovery of tracrRNA and its role in crRNA maturation

In 2011, **Emmanuelle Charpentier** and colleagues reported on the mechanisms of crRNA maturation in *S. pyogenes* [27]. Using differential RNA sequencing to characterize small, non-coding RNA molecules, they identified an active CRISPR locus, based on expression of pre-crRNA and mature crRNA molecules. Unexpectedly, the sequencing efforts also identified an abundant RNA species transcribed from a region 210 bp upstream of the CRISPR locus, on the opposite strand of the CRISPR array (Figure 2a).



**Figure 2.** Identification of tracrRNA in *S. pyogenes* as reported in [27]. **a.** Differential RNA sequencing (dRNA-seq) reveals expression of tracrRNA and crRNAs. Sequence reads of cDNA libraries of RNA are shown on top. Below is the genomic organisation of tracrRNA and CRISPR01/Cas loci. Red bar: tracrRNA is encoded on the minus strand and detected as 171-, 89- and ~75-nt tracrRNA species. Black rectangle inside the red bar: 36-nt sequence stretch complementary to CRISPR01 repeat. The pre-crRNA is encoded on the plus strand. Black rectangles: CRISPR01 repeats; green diamonds: CRISPR01 spacers; 511, 66 and 39-42 nt: pre-crRNA and processed crRNAs. **b.** Base-pairing of tracrRNA with a CRISPR01 repeat is represented. Cleavages observed by dRNA-seq and leading to the formation of short overhangs at the 3' ends of the processed RNAs are indicated by two black triangles.

The transcript was denoted *trans*-encoded small RNA (tracrRNA) and contained a stretch of 25 nucleotides (nt) with almost perfect complementarity (1-nt mismatch) to the repeat regions of the CRISPR locus, thus predicting base pairing with pre-crRNA [27]. The RNA duplex region that would form included processing sites for both pre-crRNA and tracrRNA, which immediately suggested that the two RNAs could be co-processed upon pairing (Figure 2b).

In support of the proposed idea, deletion of the tracrRNA locus prevented pre-crRNA processing and vice versa. **Charpentier** and colleagues also noted that a co-processed duplex involving tracrRNA and pre-crRNA would have short 3' overhangs, similar to those produced by the endoribonuclease RNase III, and they went on to demonstrate that this enzyme could process a heteroduplex formed between tracrRNA and pre-crRNA *in vitro* and was required for tracrRNA and pre-crRNA processing *in vivo*. Finally, the researchers found that processing also involved the Cas9 protein, since deletion of the *cas9* gene in bacteria impaired both tracrRNA and pre-crRNA processing. Based on their findings, **Charpentier** and coworkers suggested that the Cas9 protein acts as a molecular anchor that facilitates base pairing between tracrRNA and pre-crRNA,



which in turn allows recognition and cleavage by the host RNase III protein [27].

Previous reports had revealed the importance of Cas9 for interference. **Charpentier** and **Jennifer A. Doudna** initiated a collaboration to investigate if crRNA could be used to direct the sequence specificity of the nuclease. In contrast to what had been hypothesised in **Charpentier's** report a year earlier, addition of crRNA to purified Cas9 could not stimulate Cas9-catalysed target DNA cleavage [27, 28].

At this point, the two scientists made a crucial discovery. Addition of tracrRNA to the *in vitro* reaction triggered Cas9 to cleave the target DNA molecule. The tracrRNA thus had two critical functions: triggering pre-crRNA processing by the enzyme RNase III and subsequently activating crRNA-guided DNA cleavage by Cas9.

In a series of *in vitro* biochemistry experiments, the researchers investigated the biochemical mechanisms of the reaction [28]. The two nuclease domains in Cas9, HNH and RuvC, were each shown to cleave one strand of target DNA. Cleavage occurred 3 bp upstream of the PAM sequence, which in *S. pyogenes* has the sequence 5'-NGG-3', with N corresponding to any of the four DNA bases. Furthermore, as predicted from previous reports, target recognition and cleavage were inhibited by mutations in the PAM sequence [20].

A peculiar aspect of PAM sequence dependence was that cleavage of double-stranded DNA was sensitive to mutations in both the complementary and non-complementary strand whereas cleavage of single-stranded DNA targets was unaffected by mutations in the PAM motif. These observations led the authors to conclude that PAM motifs may be required to allow duplex unwinding [28].

Similar findings were also published in another report using the related CRISPR-Cas system in *Streptococcus thermophilus*. As in **Charpentier** and **Doudna's** work, this report also demonstrated that Cas9 cleaves within the protospacer, 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 [29].

In their study, **Charpentier**, **Doudna** and colleagues also worked to delineate the regions of tracrRNA and crRNA that are absolutely required for Cas9-catalysed cleavage of target DNA. This led to the identification of an activating domain in tracrRNA and the realisation that a "seed region" of ~10 nt in the PAM-proximal region of the target strand was especially important for target recognition.

Based on their *in vitro* biochemical analysis, the authors hypothesized that the structural features in the two RNA molecules required for Cas9-catalysed DNA cleavage could be captured in a single RNA molecule. In a crucial experiment, they demonstrated that this was indeed possible: the RNA components (crRNA and tracrRNA) of the Cas9 complex could be fused together to form an active, chimeric single-guide RNA molecule (sgRNA).

Furthermore, **Charpentier** and **Doudna** demonstrated that the sequence of the chimeric sgRNA could be changed so that CRISPR-Cas9 would target DNA sequences of interest, with the only constraint being the presence of a PAM sequence adjacent to the targeted DNA. They had

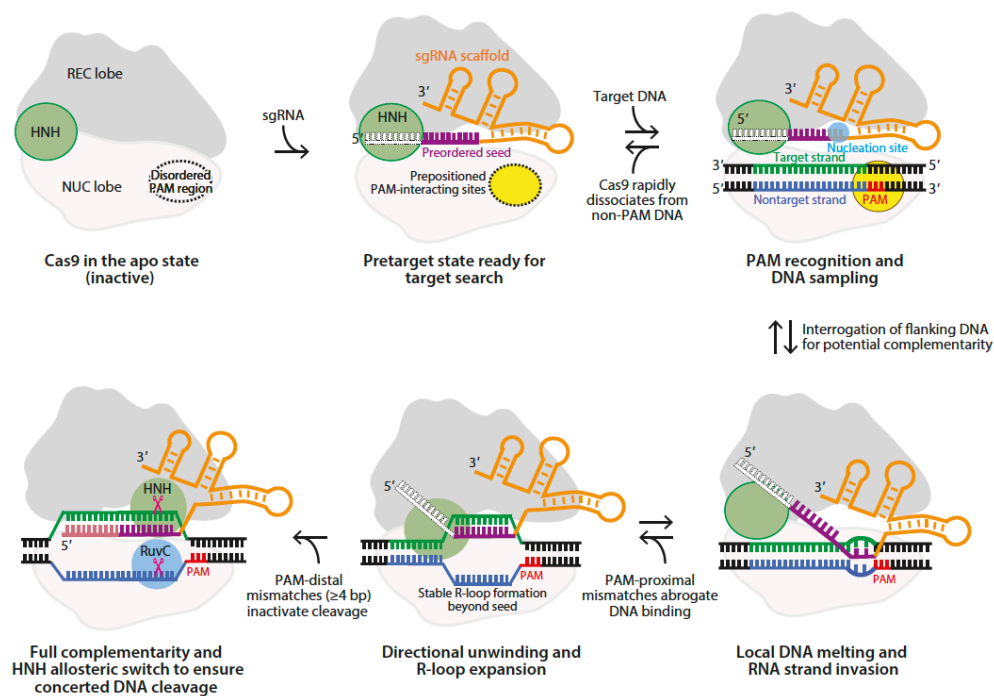
thus created a simple two-component endonuclease, containing sgRNA and Cas9, that could be programmed to cleave DNA sequences at will.

The importance of this finding was not lost on them. In the abstract of the paper reporting their findings, 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” [28].

### A molecular understanding of the CRISPR mechanism

Today, there is a detailed structural understanding of how the Cas9-gRNA complex recognizes its target and mediates cleavage. This information has been important for efforts to engineer new versions of the system, with altered PAM specificity and reduced off-target activities [30].

The structure of Cas9 in free form revealed two distinct lobes, the recognition (REC) lobe and the nuclease (NUC) lobe, with the latter containing the HNH and RuvC nuclease domains. When Cas9 binds to sgRNA, it undergoes a structural rearrangement, with the REC lobe moving towards the HNH domain (Figure 3).



**Figure 3.** A schematic representation of the mechanism by which CRISPR-Cas9 recognizes and targets DNA for cleavage as presented in [30]. Binding of sgRNA leads to a large conformational change in Cas9. In this activated conformation, the PAM-interacting cleft (dotted circle), becomes pre-structured for PAM sampling, and the seed sequence of sgRNA is positioned to interrogate adjacent DNA for complementarity to sgRNA. The process starts with PAM recognition, which in the next step leads to local DNA melting and RNA strand invasion. There is a step-wise elongation of the R-loop formation and a conformational change in the HNH domain to ensure concerted DNA cleavage. Abbreviations: bp, base pair; NUC, nuclease lobe; PAM, protospacer adjacent motif; REC, recognition lobe; sgRNA, single-guide RNA.





For target recognition, the 20-nt spacer sequence must form complementary base pairs with the protospacer sequence. In the structure of Cas9 in complex with sgRNA, the 10-nt seed sequence in the spacer adopts an A-form conformation and is positioned to engage with the target sequence in DNA [31, 32]. The seed sequence is located in the 3' end of the 20-nt spacer sequence and is essential for target recognition [25, 28, 33]. In genome editing, similarities between the seed sequence and genome sequences can cause off-target effects, even if there are many mismatches elsewhere in the spacer region of sgRNA [34].

As noted, a PAM sequence must also be present next to the target site, and mutations in this motif prevent Cas9-dependent cleavage at the target sequence. The Cas9 protein first searches for the PAM sequence, and once found, probes the flanking DNA for complementarity to the sgRNA. The GG dinucleotides in PAM are recognized by base-specific hydrogen-bonding interactions with two arginine residues in a PAM interacting site, which is disordered in the apo-form of Cas9, but becomes ordered after sgRNA binding. The interactions between PAM and Cas9/sgRNA lead to destabilization of the adjacent double-stranded DNA, which in turn facilitates for sgRNA to invade the double-stranded DNA. The destabilization is in part explained by a kink in the target DNA strand, which is caused by Cas9 interactions with the phosphate group immediately upstream of the PAM in the same strand [22].

Once a stable RNA–DNA duplex, an R-loop, has been formed, Cas9 is activated for DNA cleavage. Each of the two nuclease domains cleaves one strand of the target double-stranded DNA at a specific site 3 bp from the 5'-NGG-3' PAM sequence, and in most cases, the ends that are formed are blunt. By inactivating one of the two domains, a nickase can be formed, i.e. an enzyme that cleaves only one strand of a DNA duplex [28, 29]. Nickases are very useful for practical applications of CRISPR-Cas systems, since they can be programmed to target opposite strands and thus make staggered cuts within the target DNA. In this way, a Cas9 nickase mutant, combined with a pair of sgRNA molecules, can introduce targeted double-strand breaks with very high sequence specificity [35].

### **The application of the CRISPR-Cas9 technology in higher cells**

Genome editing relies on the existence of natural pathways for DNA repair and recombination. Double-stranded breaks typically lead to either non-homologous end joining (NHEJ) repair or homology-directed repair (HDR). In the case of NHEJ, the ends are directly ligated back together and the process usually results in a small insertion or deletion of DNA at the break, frequently causing frame shifts in coding sequences and loss of protein expression. The HDR pathway instead uses a homologous DNA sequence as a template to repair the break. By introducing modified genetic sequences as templates for the HDR, it is thus possible to introduce defined genomic changes such as base substitutions or insertions.

DNA can be introduced into mice embryonic stem cells and recombine there with the matching sequence within the host genome to produce gene-modified animals. This method is powerful but labour-intensive, since recombination events are rare and require a selectable marker, such as an antibiotic resistance gene, to be identified. Recombination efficiency is enhanced if a double-stranded break is introduced at the site of the desired recombination event, which led to a search for endonucleases that can be programmed to cleave DNA at locations of interest.

An important earlier step in the engineering of sequence-specific nucleases came with the development of zinc finger nucleases (ZFNs) and transcription activator–like effector nucleases



(TALENs). When linked to a nuclease domain, zinc finger proteins can function as site-specific nucleases that can cleave genomic DNA in a sequence-specific manner and stimulate site-specific recombination [36, 37]. TALENs provide yet another DNA-binding modality that recognizes DNA in a modular fashion and that can be fused to nuclease domain [38]. Both ZFNs and TALENs are powerful tools for genome editing. However, their widespread use has been limited by the inherent difficulties of protein design, synthesis and validation.

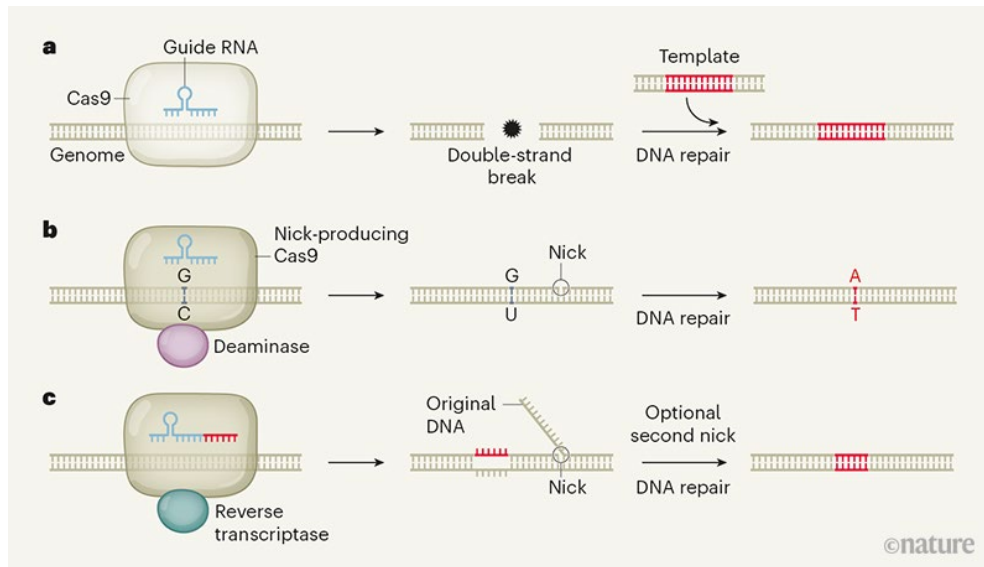
In their work, **Charpentier** and **Doudna** defined a simple two-component system that could rapidly be programmed for sequence-specific cleavage of target DNA and thereby sparked a revolution in genome editing. The first experimental demonstration that CRISPR-Cas9 could indeed be harnessed for genome editing in human and mouse cells came in early 2013 [39, 40]. These influential studies demonstrated that Cas9 nucleases could be directed by crRNA of a defined sequence to induce precise cleavage at endogenous genomic loci in mouse and human cells. For the reaction to occur, tracrRNA, crRNA, and Cas9 were all required, whereas RNase III was replaced by endogenous enzyme activities.

Just as observed by **Charpentier** and **Doudna** *in vitro*, the system could be further simplified *in vivo*, and a chimeric sgRNA molecule together with Cas9 was sufficient to cleave target DNA. The system has also been used to introduce genome modification in a number of other eukaryotic systems [41], including *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Danio rerio* and *Arabidopsis thaliana* [42-46], demonstrating its broad applicability.

In ongoing work, scientists are trying to expand the usefulness of the CRISPR-Cas system for genome editing. In addition to Cas9 from *S. pyogenes*, a number of other Cas homologues are used today for genome editing and related purposes. Naturally occurring CRISPR systems have other PAM requirements, and in addition, new Cas9 variants are continually engineered to have altered PAM compatibilities. CRISPR-Cas systems can also be used to target RNA. Studies of *Pyrococcus furiosus* demonstrated that in this species, the system encodes for a crRNA-guided Cas complex, which targets foreign mRNA [47].

Efforts are also under way to develop evermore precise CRISPR-Cas-based genome editing strategies [48]. These efforts include strategies for base editing at specific sites in eukaryotic genomes (Figure 4). As an example, a cytidine deaminase enzyme has been fused to a mutant form of Cas9 that cleaves only one strand – a nickase. When programmed with sgRNA for the desired sequence, this system can be targeted to a specific genomic location, induce a nick in the DNA there, and mediate the direct conversion of cytidine to uridine, which after replication results in a cytosine-to-thymine conversion [49].

Another elegant example is a method called prime editing, in which a Cas9 nickase is fused to a reverse transcriptase enzyme [50]. In this approach, the sgRNA contains an additional piece of RNA, creating a “prime editing guide RNA” that both specifies the target site and encodes the desired edit. Once produced by the reverse transcriptase, the DNA synthesized can be installed at the nick, replacing one of the original DNA sequences.



**Figure 4.** Genome editing with Cas9 as presented in [48]. **a.** The Cas9 enzyme is directed to target DNA by a guide RNA and produces a double-stranded break. A piece of DNA can be used as a template for homology-directed repair (HDR). **b.** Cas9 can be fused to a deaminase enzyme. The mutant Cas9 produces a nick, which stimulates deaminase activity. The deaminase converts a cytidine base (C) to uracil (U). DNA repair then repairs the nick and converts a guanine–uracil (G–U) intermediate to an adenine–thymine (A–T) base pair. **c.** Prime editing. A nick-producing Cas9 and a reverse transcriptase enzyme produce nicked DNA, into which sequences corresponding to the guide RNA have been incorporated. The original DNA sequence is cut off, and DNA repair then fixes the nicked strand to produce a fully edited duplex.

## Concluding remarks

In 2012, **Charpentier** and **Doudna** reported “that the Cas9 endonuclease can be programmed with guide RNA engineered as a single transcript to cleave any double-stranded DNA sequence” [28]. Their discovery has led to widespread applications of the CRISPR-Cas9 system as a powerful and versatile tool in genome editing.

By introducing a vector encoding the Cas9 nuclease and an engineered sgRNA, scientists are now able to make precise single-base-pair changes or larger insertions. Coupled with the availability of genome sequences for a growing number of organisms, the technology allows researchers to explore these genomes to find out what genes do, move mutations that are identified as associated with disease into systems where they can be studied and tested for treatment, or where they can be tested in combinations with other mutations. The technology has enabled efficient targeted modification of crops and is currently being developed to treat and cure genetic diseases, for instance by modifying hematopoietic stem cells to treat sickle cell disease and  $\beta$ -thalassemia.

Finally, it should be emphasised that the power of the CRISPR-Cas9 technology also raises serious ethical and societal issues. It is of utmost importance that the technology is carefully regulated and used in responsible manner. To this end, the World Health Organization has recently established a global multi-disciplinary expert panel to examine the scientific, ethical, social and



legal challenges associated with human genome editing, with the aim to develop a global governance framework for human genome editing.

*Claes Gustafsson*

Professor of Medical Chemistry  
Member of the Royal Swedish Academy of Sciences  
Member of the Nobel Committee for Chemistry

## References

1. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
2. Knott, G.J. and J.A. Doudna, *CRISPR-Cas guides the future of genetic engineering*. Science, 2018. **361**(6405): p. 866-869.
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, 1987. **169**(12): p. 5429-33.
4. Mojica, F.J., G. Juez, and F. Rodriguez-Valera, *Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites*. Mol Microbiol, 1993. **9**(3): p. 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, 2000. **36**(1): p. 244-6.
6. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes*. Mol Microbiol, 2002. **43**(6): p. 1565-75.
7. Haft, D.H., et al., *A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes*. PLoS Comput Biol, 2005. **1**(6): p. e60.
8. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action*. Biol Direct, 2006. **1**: p. 7.
9. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements*. J Mol Evol, 2005. **60**(2): p. 174-82.
10. Pourcel, C., G. Salvignol, and G. Vergnaud, *CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies*. Microbiology (Reading), 2005. **151**(Pt 3): p. 653-663.
11. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin*. Microbiology (Reading), 2005. **151**(Pt 8): p. 2551-2561.
12. Tang, T.H., et al., *Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus*. Proc Natl Acad Sci U S A, 2002. **99**(11): p. 7536-41.



13. Makarova, K.S., et al., *Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants*. Nat Rev Microbiol, 2020. **18**(2): p. 67-83.
14. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
15. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science, 2008. **321**(5891): p. 960-4.
16. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA*. Science, 2008. **322**(5909): p. 1843-5.
17. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA*. Nature, 2010. **468**(7320): p. 67-71.
18. Horvath, P., et al., *Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus*. J Bacteriol, 2008. **190**(4): p. 1401-12.
19. Mojica, F.J.M., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system*. Microbiology (Reading), 2009. **155**(Pt 3): p. 733-740.
20. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus*. J Bacteriol, 2008. **190**(4): p. 1390-400.
21. Wang, J., et al., *Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems*. Cell, 2015. **163**(4): p. 840-53.
22. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. Nature, 2014. **513**(7519): p. 569-73.
23. 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, 2011. **45**: p. 273-97.
24. Terns, M.P. and R.M. Terns, *CRISPR-based adaptive immune systems*. Curr Opin Microbiol, 2011. **14**(3): p. 321-7.
25. Sapranauskas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli*. Nucleic Acids Res, 2011. **39**(21): p. 9275-82.
26. Hille, F., et al., *The Biology of CRISPR-Cas: Backward and Forward*. Cell, 2018. **172**(6): p. 1239-1259.
27. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III*. Nature, 2011. **471**(7340): p. 602-7.
28. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
29. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
30. Jiang, F. and J.A. Doudna, *CRISPR-Cas9 Structures and Mechanisms*. Annu Rev Biophys, 2017. **46**: p. 505-529.
31. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation*. Science, 2014. **343**(6176): p. 1247997.
32. Jiang, F., et al., *STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition*. Science, 2015. **348**(6242): p. 1477-81.
33. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9*. Nature, 2014. **507**(7490): p. 62-7.



34. Pattanayak, V., et al., *High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity*. Nat Biotechnol, 2013. **31**(9): p. 839-43.
35. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity*. Cell, 2013. **154**(6): p. 1380-9.
36. Beumer, K., et al., *Efficient gene targeting in Drosophila with zinc-finger nucleases*. Genetics, 2006. **172**(4): p. 2391-403.
37. Moehle, E.A., et al., *Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases*. Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3055-60.
38. Christian, M., et al., *Targeting DNA double-strand breaks with TAL effector nucleases*. Genetics, 2010. **186**(2): p. 757-61.
39. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems*. Science, 2013. **339**(6121): p. 819-23.
40. Mali, P., et al., *RNA-guided human genome engineering via Cas9*. Science, 2013. **339**(6121): p. 823-6.
41. Mali, P., K.M. Esvelt, and G.M. Church, *Cas9 as a versatile tool for engineering biology*. Nat Methods, 2013. **10**(10): p. 957-63.
42. DiCarlo, J.E., et al., *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems*. Nucleic Acids Res, 2013. **41**(7): p. 4336-43.
43. Gratz, S.J., et al., *Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease*. Genetics, 2013. **194**(4): p. 1029-35.
44. Friedland, A.E., et al., *Heritable genome editing in C. elegans via a CRISPR-Cas9 system*. Nat Methods, 2013. **10**(8): p. 741-3.
45. Hwang, W.Y., et al., *Efficient genome editing in zebrafish using a CRISPR-Cas system*. Nat Biotechnol, 2013. **31**(3): p. 227-9.
46. Li, J.F., et al., *Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9*. Nat Biotechnol, 2013. **31**(8): p. 688-91.
47. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex*. Cell, 2009. **139**(5): p. 945-56.
48. Platt, R.J., *CRISPR tool modifies genes precisely by copying RNA into the genome*. Nature, 2019. **576**(7785): p. 48-49.
49. Komor, A.C., et al., *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage*. Nature, 2016. **533**(7603): p. 420-4.
50. Anzalone, A.V., et al., *Search-and-replace genome editing without double-strand breaks or donor DNA*. Nature, 2019. **576**(7785): p. 149-157.