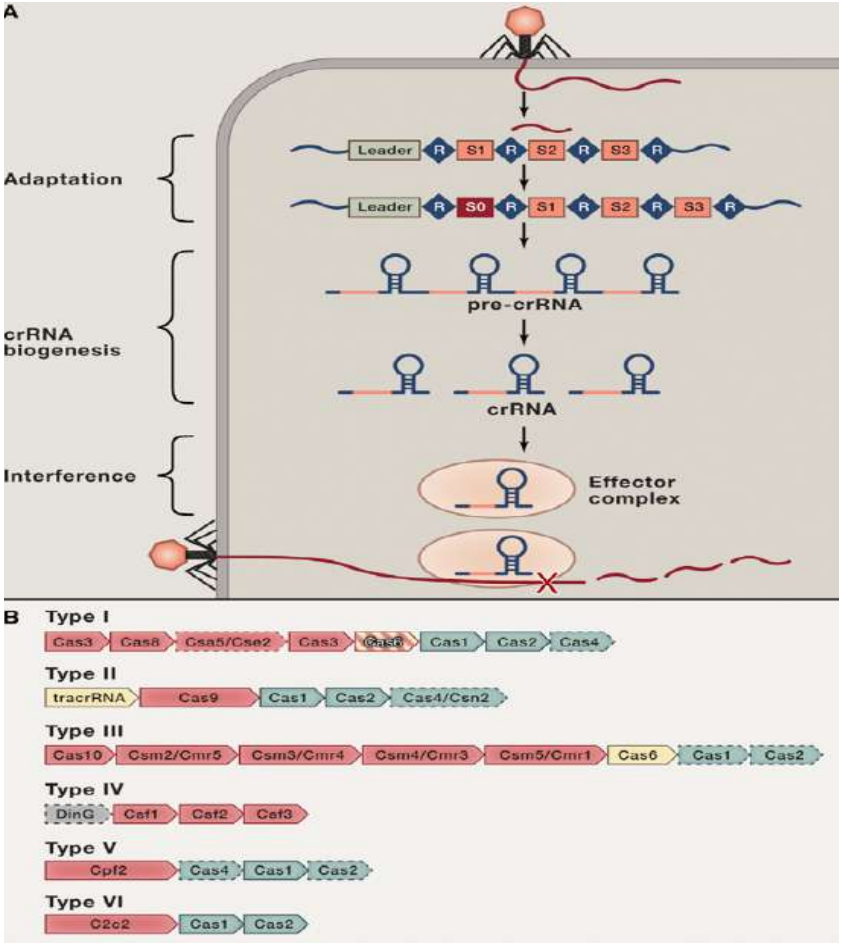
CRISPR-Cas9: Revolution In Gene Editing

Mou Saha ; Sreayosi Sinha Roy

[1]Assistant professor ; Department of Microbiology, R.K.V.M. Sarada Ma Girls' College ; Barasat, Kolkata-700126; Email- saha.mou136@gmail.com

[2] Post Graduate Student of Biomedical Genetics; Vellore Institute Of Technology, Vellore; Email- sreayosis@gmail.com

Gene editing and CRISPR Cas-9



Genome editing provides scientists with the capability to manipulate the DNA of organisms and thus encompasses an innovative line-up of technologies. These innovative techniques empower researchers to modify, insert or remove genetic material at exact locations within an organism's genome. Among these advancements, one notably prominent technique is CRISPR-Cas9 , denoting "clustered regularly interspaced short palindromic repeats" and "CRISPR-associated protein 9" is one of the most prominent techniques. Within the scientific community, compared to other methods of gene editing CRISPR-Cas9 has incited significant enthusiasm due to its remarkable attributes, including its swiftness, affordability, precision, and efficiency.

Fig: organization of CRISPR-Cas system
https://www.cell.com/cell/fulltext/S0092-8674(15)01692?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867415016992%3Fshowall%3Dtrue

Using CRISPR-Cas9 as a form of immune defence, a naturally occurring genome editing mechanism, bacteria capture fragments of viral DNA and integrate them into their own DNA in a precise pattern, resulting in CRISPR arrays.

Arrays such as these enable bacteria to "recall" viruses or closely related organisms. If a subsequent viral attack occurs, the bacteria produce RNA segments from the CRISPR arrays that attach to certain segments of the viral DNA. In this regard, bacteria use enzymes such as Cas9 to cleave the viral DNA, thereby neutralizing the threat posed by the virus.

○ History

- **The invention of clustered DNA repeats** - Three different regions of the world separately made the finding of clustered DNA repeats. Researcher Yoshizumi Ishino and his colleagues at Osaka University first described the technology that would later be known as CRISPR in 1987. They unintentionally cloned the "iap" gene (isozyme conversion of alkaline phosphatase), which was their aim, along with a portion of a CRISPR sequence. The repeats were organised in an innovative way. Repeated sequences are often organised in a straight line without other sequences in between. They were unaware of the purpose of the broken clustered repeats.A cluster of interrupted direct repeats (DR) in Mycobacterium TB was the subject of a pair of studies written by Dutch researchers in 1993. They identified the diversity of the sequences that stood between the direct repeats in various M. tuberculosis strains and used this characteristic to create the spoligotyping method, which is still in use today.

  At the University of Alicante in Spain, Francisco Mojica studied the purpose of repeats in archaeal organisms like Haloferax and Haloarcula. He believed clustered repeats were responsible for segregating duplicated DNA into daughter cells during cell division. For the first time, transcription of the interrupted repetitions was also identified and this was **CRISPR's first comprehensive characterization.** They determined that 20 species of microorganisms with interrupted repeats belonged to the same family. Mojica initially referred to these sequences as "short regularly spaced repeats" (SRSR) because of the interspaced nature of those sequences. The acronym **CRISPR** (Clustered Regularly Interspaced Short Palindromic Repeats) was proposed in 2001 by **Mojica** and **Ruud Jansen**, who were looking for further interrupted repeats in order to reduce the confusion caused by the multiple acronyms used to describe the sequences in the scientific literature.

- **Cas9 and PAM region Identification**: Alexander Bolotin , a researcher of National Institute for Agricultural Research, France, was researching the recently sequenced bacteria Streptococcus thermophilus, which featured a unique CRISPR locus. The CRISPR array resembled previously described systems, but several of the known cas

genes were missing. Instead, it contained unique cas genes, including one that encoded Cas9, a big protein that they projected to have nuclease activity. They also observed that the spacers, having homology to viral genes share a common sequence at one end. He classified these sequences as **The protospacer adjacent motif (PAM).**

- **Hypothetical approach to adaptive immunity:** Eugene Koonin, a researcher of the US National Centre for Biotechnology Information**,** dismissed the earlier theory that the Cas proteins might make up a novel DNA repair system while researching clusters of orthologous protein groups using computational analysis. He then hypothetically proposed CRISPR cascades as a part of bacterial immune system depending on insertion of homologous DNA to phage DNA in the portion of spacer array.

- **Adaptive immunity demonstrated experimentally:** *S. thermophilus* is a popular strain used in the dairy sector to produce yoghurt and cheese. Researchers at Danisco wanted to learn more about how S. thermophilus handles phage attack, a typical issue in the production of industrial yoghurt. According to an experimental demonstration by Horvath and colleagues, CRISPR systems do actually function as an adaptive immune system since they incorporate fresh phage DNA into their array. Additionally, they demonstrated that Cas9 is most likely the only protein necessary for interference, the CRISPR system's method of neutralising invading phage was yet unknown.

- **Transcription of spacer sequences into guide RNAs:** in 2008, John van der Oost, a scientist of the University of Wageningen, Netherlands highlighted some of the specifics of how CRISPR-Cas systems "interfere" with invading phage were being brought in by scientists. The first important piece of information was provided by John van der Oost and colleagues, who demonstrated that in *Escherichia coli*, spacer sequences obtained from phage are translated into short RNAs known as CRISPR RNAs (crRNAs), which direct Cas proteins to the target DNA.

- **Discovery of the target molecule**: Marraffini and Sontheimer's invention was the next crucial step in understanding the mechanism of interference (Marraffini and Sontheimer, 2008) that not RNA but DNA is the target molecule. This invention was surprizing since mostly everyone believed that CRISPR had similarities with eukaryotic RNAi silencing mechanisms, targeting the RNA. In their study, Sontheimer and Marraffini stated that if this method gets applied to non-bacterial systems, it will behave as a very useful tool.

- **Cleaving mechanism of Cas9**: In 2010, Sylvain Moineau, a researcher of University of Laval, Canada and his colleagues showed that CRISPR-Cas9 causes double-stranded breaks in the target DNA Additionally, they verified that the CRISPR-Cas9 system only requires Cas9 as the necessary protein for cleavage. This is a characteristic of Type II

CRISPR systems, in which interference is carried out by a solitary big protein (in this case, Cas9) working with crRNAs.

- **tracrRNA discovery**: In 2011, The team of Emmanuelle Charpentier provided the ultimate understanding of the mechanism of natural CRISPR-Cas9-guided interference. They implemented small RNA sequencing on the CRISPR-Cas9-equipped Streptococcus pyogenes. As a result , they found a second short RNA, which was then named trans-activating CRISPR RNA (tracrRNA), in addition to the crRNA. In conclusion, they stated that Cas9 is guided to its targets by the duplex created by tracrRNA and crRNA.

- **Biochemical characterization of Cas9 mediated cleavage**: Researcher Virginijus Siksnys and his fellows examined E. coli strain to purify Cas9 in complex with crRNA. They modified it with CRISPR locus from the S. thermophilus and performed different biochemical experiments to understand the mode of action. This was done to take advantage of their heterologous system. Scientists confirmed the cleavage location and the need for the PAM . They used point mutations they proved that HNH domain present in Cas9 cleaves the complementary site. On the other hand, non-complementary strands are ckeaved by the RuvC domain only. Additionally, they mentioned that the crRNA may be reduced to a 20-nt stretch that would be sufficient for effective cleavage. Most impressively, they demonstrated that by altering the sequence of the crRNA, they could programme Cas9 to target a specific location.

  Simultaneously, Charpentier and Jennifer Doudna, researchers of University of California, Berkeley demonstrated that the crRNA and the tracrRNA could be joined together to provide a single, synthetic guide.

  - **Introducing CRISPR Cas9 into genetic engineering**: In 2013, American researcher Feng Zhang, used two Cas9 orthologs derived from S. pyogenes and S. thermophilus. Zhang and his team performed targeted genome cleavage in mice and human cells. Their experiment led to the first successful adaptation of CRISPR-Cas9 for genome editing in eukaryotic cells. Additionally, they illustrated how the first system was assumed to be set up to target numerous genomic locations, and the second system could activate homology-directed repair. Researchers from George Church's Harvard University lab published similar findings in the same issue of Science.

○ Structural configurations of CRISPR Cas9

Two main components of CRISPR Cas9 are there in the prokaryotic system:

→ CRISPR arrays and

→ CRISPR-associated (Cas) proteins.

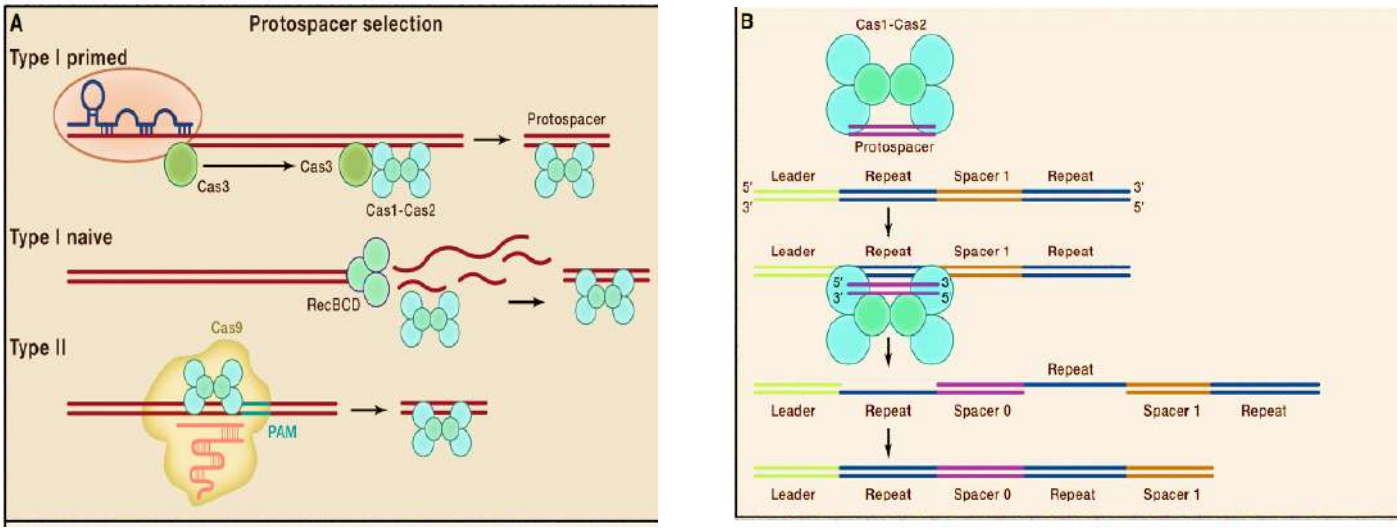


Fig: Process of Protospacer selection and insertion as Spacer
https://www.cell.com/cell/fulltext/S002-8674(15)01692?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867415016992%3Fshowall%3Dtrue

The first component is termed as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) which is a unique type of repetitive DNA stretches. This This CRISPR array is all in one part of the chromosome, gathering in one cluster.

- Regularly interspaced is referring the spaces that are regularly placed between these repeats all along the CRISPR region. **Spacers** are highly variable sequences that comprise the prokaryotes 'immunological memory
  Generation of spacers: when a bacteriophage injects its DNA inside a bacteria, a pair of enzymes (Cas1 and Cas2) works together and cuts off a region of viral DNA. This piece of viral DNA is called a **protospacer**. This protospacer is then get inserted into the 5'end of the CRISPR array as a **spacer.** In this way, the protospacers get inserted into the array as spacers eventually build a new repeat region.

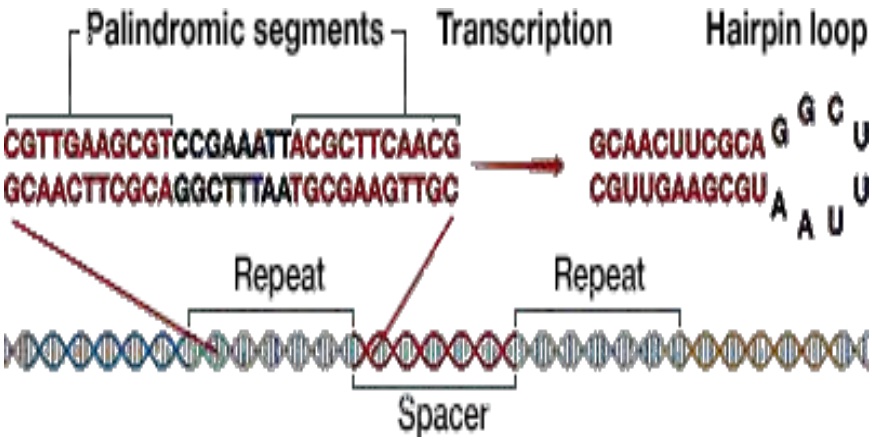


Fig: Parts of CRISPE array
https://www.cell.com/cell/fulltext/S002-8674(15)01692?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867415016992%3Fshowall%3Dtrue

- Another distinctive and essential element within the CRISPR array is **palindromic repeats**. In other words, these are regions in which a DNA or RNA sequence reads the same both forward and backwards. They play a fundamental role in the system's functioning. These palindromic repeats provide a distinctive pattern that is translated into RNA in conjunction with the spacer sequences. The Cas proteins, in particular the Cas9 enzyme, are directed by these RNA molecules, known as "guide RNAs" (gRNAs) or "crRNAs," to recognise and target particular viral DNA sequences for elimination. The complementary nature of the spacer sequence in the guide RNA and the target virus DNA is what enables identification. The presence of these repeats emphasises how the CRISPR-Cas system is supported by structural components that are conserved across diverse bacterial species.

- A crucial part of the CRISPR-Cas9 system is the **PAM region**. It ensures that only the other DNA sequences in the genome cuts off by the Cas9 protein. Fundamentally this region ensures the specificity of the CRISPR-Cas9 system. The **protospacer adjacent motif**, or **PAM region**, is a short DNA sequence that is present right next to the target DNA sequence that the CRISPR-Cas9 system will cut. Normally this region consists of 2–6 base pairs and is palindromic, which means that it may be read both forward and backwards. The sequence 5'-NGG-3', where "N" can be any nucleotide, is the most frequent PAM sequence. However, other PAM sequences, including 5'-NG-3' and 5'-NAG-3', are also feasible.

  The two steps of the CRISPR-Cas9 system are **adaptation** and **interference**. The interference stage is responsible for the identification and chopping of invasive DNA, while the adaption stage is responsible for acquiring new spacers from the invasive DNA. The PAM sequence is necessary for both the adaptation and interference processes, but the degree to which it is recognised differs for each stage. This is so because the steps of adaptation and interference use various molecular processes.

  PAM elements have been separated into target interference motifs (TIMs) and spacer acquisition motifs (SAMs) to accommodate this variation. The functional motif connected to the protospacer that is recognised by the spacer acquisition machinery before the excision is known as a SAM element. The in-depth knowledge about CRISPR Cas-9 and gene editing is solely based on the understanding of Pam regions.

- The CRISPR-Cas9 system exploits pre-CRISPR RNA (pre-crRNA) as a precursor to CRISPR RNA (crRNA), a short RNA molecule that targets and cleaves DNA. **Pre-CRISPR RNA** and **Cas mRNA** are generated during the transcription of the CRISPR and CAS genes, respectively. The mature crRNA is subsequently created by processing the **pre-crRNA**. The protospacer sequence is complementary to the short RNA molecule known as the mature **crRNA.** A ribonucleoprotein complex formed by the

mature crRNA and the translated Cas protein targets and cleaves the viral DNA in a sequence-specific way.

- Another part of the CRISPR-Cas9 system is **trans-activating CRISPR RNA (tracrRNA**). These are short RNA molecule that aids in directing the Cas9 protein to the target DNA sequence. The tracrRNA actually "trans-activates" the CRISPR-Cas9 system giving rise to the concept, which is if it is present in the same cell as the crRNA and Cas9 protein, it can activate the CRISPR-Cas9 system. These tracer RNAs have complementary repeats in them to regions in the CRISPR RNA.

- An enzyme RNAase comes along then simultaneously and cuts through those repeated regions. This action results in producing pieces of RNAs made up of spacers, repeats and tracer RNA. A structure formation takes place within the cell containing

  → one piece of spacer RNA and repeat RNA- which acts as a single polymer of nucleotide and
  → tracrRNA – another polymer of RNA nucleotide held to the previous polymer by hydrogen bonds.

  This structure is called **cr:tracrRNA.** This molecule is then get picked up by an enzyme called Cas-9. After binding of Cas-9, cr:tracrRNA is called **gRNA** or **guide RNA**. The Cas proteins then cut the foreign DNA at a specific location.

Cas9

A 160 kilodalton protein known as Cas9 (CRISPR-associated protein 9, formerly Cas5, Csn1, or Csx12) is essential for certain bacteria's immune defence against DNA viruses and plasmids. Cas9 is an RNA dependent DNA endonuclease thereby having the ability to cut DNA. A guide RNA (gRNA), a kind of RNA, directs Cas9 to a certain DNA sequence. The HNH (Hen1-NPH2) domain of the Cas9 protein allows it to recognise the PAM region. The Cas9 protein positions itself for cleavage of the target DNA sequence by the HNH domain, which binds to the PAM region. Cas9 breaks the DNA at a specific region known as the protospacer adjacent motif (PAM) when it binds to the gRNA.

○ STRUCTURAL CONFIGURATION OF CAS-9

Two Cas9-sgRNA-DNA ternary complexes (Mol A and Mol B) were present in the crystallographic asymmetric unit. The sgRNA and the DNA are both recognised by Cas9 in comparable ways despite the two complexes having different conformations. The HNH

domain in Mol B is flexible, in contrast to Mol A where it is coupled to the RuvC domain via a disordered linker. Therefore it is important to describe the structural characteristics of Mol A before talking about the structural variations between the two complexes, which is responsible for the conformational flexibility of Cas9.
Cas9 has two lobes: a **recognition (REC) lobe** and a **nuclease (NUC) lobe**, as shown by the crystal structure

REC Lobe: The CRISPR-Cas9 protein has a domain called the REC lobe that is in charge of binding to the guide RNA. The REC lobe is classified into three regions:

→ the REC1 region or domain
→ he REC2 domain, and
→ a lengthy helix, known as the bridge helix.

A new study revealed that the **REC lobe** has no structural similarities to other well-known proteins, proving that it is a functional domain unique to the Cas9 gene. REC lobe is one of the least conserved sections in the type II CRISPR system's three Cas9 families (IIA, IIB, and IIC), and many Cas9 orthologs have REC lobes that are noticeably shorter.

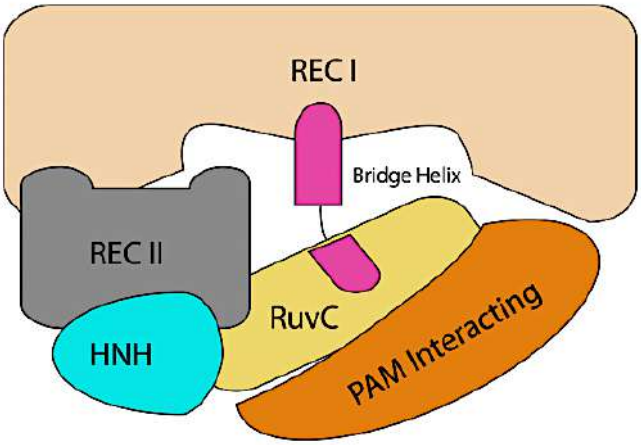


Fig: Different regions of REC lobe

REC1 Domain: The Cas9 protein's REC1 domain, which is the biggest domain, is in charge of binding to the guide RNA. Through a sequence of interactions, the amino acid residues of the protein and the guide RNA's nucleotides accomplish this. The Cas9 protein and the guide RNA are positioned for DNA cleavage by the REC1 domain. A number of positively charged residues in the REC1 domain specifically bind with the negatively charged phosphate backbone of the guide RNA. The hydrophobic bases of the guide RNA can be interacted with by a variety of the hydrophobic residues that are present in it. The Cas9 protein and the guide RNA combine to create a stable complex as a result of these interactions.

REC2 Domain: The CRISPR-Cas9 protein has a little domain called the REC2 domain, which is found in the REC lobe. The interaction between the Cas9 protein and the guide RNA is assumed to be stabilised by the REC2 domain. It accomplishes this by interacting with the bridge helix and the REC1 domain. It is also believed that the regulation of Cas9 activity involves the REC2 domain. Studies have demonstrated that changes to the REC2 domain can cause Cas9 activity to decline. Although the REC2 domain's precise purpose is not yet fully understood, it is believed to be a crucial component of the CRISPR-Cas9 system. The following are some potential REC2 domain uses:

→ preventing the Cas9 enzyme and guide RNA from forming an unstable complex
→ control of Cas9 activity
→ interacting with other CRISPR-Cas9 system proteins

Bridge helix (BH): The CRISPR-Cas9 protein has a short, arginine-rich helix called the bridge helix (BH) that connects the REC and NUC lobes. It is believed to be crucial for directing the Cas9 enzyme and the guide RNA where to cleave the DNA. Six positively charged amino acid residues called arginine residues make up the BH. These arginine residues bind with the guide RNA's negatively charged phosphate backbone, stabilising the Cas9 protein and guide RNA combination. As it connects the NUC and REC lobes and interacts directly and indirectly with target DNA, tracrRNA, and crRNA, BH plays a crucial part in the function.Cas9 activity can be reduced by mutating the arginine residues in the BH. This is due to the fact that the BH is necessary for both stabilising the Cas9 protein and guide RNA complex and for placing the Cas9 protein and guide RNA appropriately for DNA cleavage.

NUC Lobe: The C-terminal domain of the CRISPR-Cas9 protein is known as the NUC lobe. The RuvC, HNH, and PAM-interacting (PI) domains make up the NUC lobe. At the junction of the REC and NUC lobes, a positively charged groove accommodates the negatively charged sgRNA: target DNA heteroduplex. The three separate RuvC motifs (RuvC I–III) that make up the RuvC domain are put together in the NUC lobe, where they interact with the PI domain to create a positively charged surface that interacts with the 3′ tail of the sgRNA. Only a few interactions exist between the HNH domain and the rest of the protein, which is located between the RuvC II and III motifs.

The complementary DNA strand to the guide RNA is split apart by the HNH domain. The DNA strand that is not complementary to the guide RNA is cut by the RuvC domain. The PAM sequence is a short DNA sequence for binding of PAM-interacting domain. It plays essential role in DNA cleavage mechanism of Cas9. NUC lobe is required for the cleavage of DNA by Cas-9 system. In case of binding to the PAM sequence, The PAM-interacting domainIS IN-CHARGE. On the other hand, cleaving the DNA strand is dependent on the HNH and RuvC nuclease domains

PAM Interacting Domain: The domain for the CRISPR-Cas9 protein binding to the PAM sequence is known as the PAM-interacting domain (PID) . The BH also engage in interaction. The PAM sequence, short DNA sequence with which the **PAM-interacting domain** binds to, is necessary for Cas9 to cleave DNA. The Cas9 protein and the guide RNA are positioned correctly for DNA cleavage thanks to the contact between the BH and the PAM-interacting domain.

The C-terminus of the Cas9 protein is where the PID is found. A number of conserved amino acid residues help to stabilise the structure's -hairpin shape. The PAM sequence and these conserved amino acid residues work together to bind the DNA-binding PAM-interacting domain. The PID is crucial to the CRISPR-Cas9 system's specificity. This is so because each CRISPR-Cas9 system has a different PAM sequence. The PID makes sure that Cas9 only cleaves DNA that has the PAM sequence by attaching to it. A decrease in Cas9 specificity can result from PID mutations. This is because binding to the PAM sequence requires the PID. Cas9 may cleave DNA that does not contain the PAM sequence if the PID is altered because it may no longer be able to connect to the PAM sequence.

HNH Domain: For the cleavage of the complementary DNA strand to the guide RNA, CRISPR-Cas9 protein has a nuclease domain called the HNH domain. It can be detected in the Cas9's C-terminus. There are two subdomains: HNH1 and HNH2 subdomains.

→ HNH1 Subdomain: The smallest of the two HNH domain subdomains is the HNH1 subdomain. It is in charge of attaching to the guide RNA's complementary DNA strand. It accomplishes this through several interactions between its amino acid residues and the DNA strand's nucleotides. A number of conserved amino residues help to stabilise the -hairpin structure of the HNH1 subdomain. There are 80 amino acid residues in it. Ten hydrogen bonds help to stabilise its -hairpin structure. The DNA strand and these conserved amino acid residues interact for attaching HNH1 subdomain to the DNA. The CRISPR-Cas9 system cannot modify DNA without the HNH1 subdomain. Cas9 may become less active or may cleave the incorrect DNA strand as a result of mutations in the HNH1 subdomain.
The DNA strand is depicted in green, while the HNH1 subdomain is displayed in blue. The DNA strand's nucleotides are depicted in black, while the HNH1 subdomain's conserved amino acid residues are highlighted in red.

→ HNH2 subdomain: Of the two subdomains of the HNH domain, the HNH2 subdomain is the larger. It is in charge of cleaving the guide RNA's corresponding DNA strand. By first creating a phosphodiester link with the DNA strand and then hydrolysing that bond, it accomplishes this. Numerous amino acid residues make up the HNH2 subdomain and are required for its proper operation. The pockets formed by these amino acid residues can hold the DNA strand. The DNA backbone can be deprotonated by the acidic residues around the pocket, which makes it simpler to rupture the link.

RuvC Domain: Numerous proteins, including the CRISPR-Cas9 protein, have the RuvC domain, a nuclease domain. It is called after the bacterial protein RuvC, which is involved in the repair of DNA double-strand breaks and bears that name. The HNH domain and the RuvC domain share structural and functional similarities. Each domain is made up of a pocket that can hold a DNA strand and several acidic residues that can deprotonate the

DNA backbone. The domains can split the DNA strand more readily as a result. The CRISPR-Cas9 system can modify DNA without the need for the RuvC domain. It can, however, be used to cause double-stranded DNA breaks that are more challenging for cells to heal. Applications like gene editing and cancer therapy may benefit from this. The following information on the RuvC domain is provided:

→ It is comparable to the HNH domain in both structure and function.
→ It exceeds the HNH domain in size.
→ It has a unique pocket architecture.
→ For cleaving the DNA strand that is not complementary to the guide RNA, it is more specific.
→ The ability to modify DNA is not required for the CRISPR-Cas9 system to function.
→ It can be utilised to produce more challenging double-stranded DNA breaks.

Chart of domains present in Cas-9 protein

Domain	Function
REC1 domain	Binds to the guide RNA
REC2 domain	Unclear
Bridge helix	Positions the Cas9 protein and the guide RNA for DNA cleavage
PAM-interacting domain	Binds to the PAM sequence
HNH domain	Cleavage the DNA strand that is complementary to the guide RNA
RuvC domain	Cleavage the DNA strand that is not complementary to the guide RNA

sgRNA: The term "single guide RNA" is abbreviated as "sgRNA." As the name indicates, sgRNA is a single RNA molecule that combines the scaffold tracrRNA sequence with the short crRNA sequence that was specifically created for it. The tracrRNA contributes in stabilising the complex and attracting the Cas9 protein, while the gRNA binds to the DNA target region.

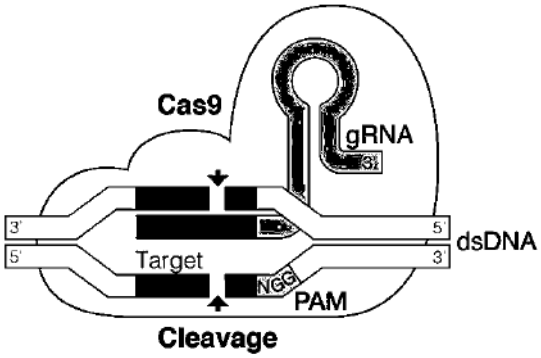


Fig: Diagram of sgRNA

The target DNA sequence and the Cas9 protein are both taken into consideration when a computer programme is used to build the sgRNA. The best sgRNA sequence is chosen from within the many generated by the programme, based on factors including binding affinity, stability, and effectiveness. According to a recent study, these algorithms are categorised into three groups: Regression models include gradient boosting regression trees and extreme gradient boosts (XGBoost), classification models include logistic regression and support vector machines, and emerging technologies or hybrid algorithms include a simple average of multiple models and deep learning technology.

The development of effective sgRNAs is becoming a more urgent task as the CRISPR/Cas9 system has spread to become a common gene editing tool in biological research. To deal with this crucial issue, numerous bioinformatics tools have been created. However, a significant drawback of earlier research is to the calibre of the datasets utilised to develop such systems. The majority of experimental techniques are based on phenotypic screening, which makes it difficult to gauge the effectiveness of CRISPR/Cas9 editing. It is uncertain that effective gene editing will result in uniform and precise phenotypic changes across all examined genes and target sites. As a result, such indirect methods add unintentional distortion to the datasets used to train machine learning algorithms, which may conceal the real properties that are unique to sgRNA-guided Cas9 cleavage. Additionally, sgRNAs studied in functional screens are often created for a limited number of genes and examined in a single cell line. These limitations may introduce biases that are particular to each experimental context, such as biases relating to varying degrees of genomic accessibility or varying reactions to DNA cleavage in a cell line- or gene-specific way. All of these elements might limit the generalizability of the model.

Classification of CRISPR-Cas system: The most recent categorization developed by Makarova et al. defines two major classes of CRISPR-Cas systems as well as six primary kinds (I-VI), which are currently divided into around 30 subtypes in total. Based on the characteristics of the protein-RNA effector complex engaged in nucleic acid-targeting, Class 1 and Class 2 systems are distinguished from one another. Class 2 effector complexes40 consist of a single Cas protein along with crRNA, occasionally joined by an extra trans-activating crRNA molecule known as tracrRNA. Class 1 effector complexes39 contain crRNA and several protein subunits, typically encoded by three to six cas genes. The first class is further divided into types I, III, and IV, while class two is divided into types II, V, and VI.

- **Cas-I System:** Type I CRISPR-Cas systems—the most prevalent in sequenced genomes—encode multi-subunit effector complexes93 as opposed to single-effector type II systems. Type I interference saw the first descriptions of numerous mechanistic findings on interference, including the existence of a seed sequence, the mechanism of R-loop creation, and inhibition by anti-CRISPRs. There are at least six different subtypes of type I systems, which all contain the distinctive protein Cas3 and make use of the DNA binding complex Cascade. In type, I systems, a mature crRNA is needed for the effector complex, which is created during pre-crRNA processing. The repeat sequences are broken up into specified stem-loop structures by a designated ribonuclease (often Cas6), resulting in a spacer sequence flanked by two repeat fragments98,99. The big subunit (Cas8) that is in charge of PAM recognition and Cas5 are both located towards the 5′ ends of the crRNA at the "base" of Cascade.

  One generalised description cannot effectively capture the variety of Type I systems, which may be the most diverse CRISPR-Cas grouping. Each subtype has

distinct properties in addition to the presence of a distinctive large subunit protein, such as the split Cas3 organisation, the functional replacement of Cas6 with Cas5 or the presence of their fusion the functional replacement of Cas8 with Cas-9, the fusion of Cas2 and Cas3, and the presence of Cas8 and Cas3. The degree of variation found in these complexes underlines the adaptability of CRISPR-Cas immune systems and illustrates the vast array of options that nature has investigated throughout the course of evolution.

- **Cas-II System:** The well-known Cas9 endonuclease, a protein originally known by the names Cas5 and Csn1, is a component of type II systems. In DNA targets that have complementarity to the 20-nucleotide crRNA guide sequence and a flanking protospacer-adjacent motif (PAM), Cas9 induces double-strand breaks (DSBs) under the guidance of a dual-RNA substrate made up of crRNA and tracrRNA. Because these sequences lack flanking PAMs, the PAM requirement ensures that Cas9 stays away from complementing targets within the CRISPR array itself. It's interesting to note that the acquisition of additional spacer sequences depends on Cas9-mediated PAM recognition. Following R-loop formation, the HNH and RuvC nuclease domains of Cas9, respectively, facilitate cleavage of the target strand (base-paired to the crRNA) and nontarget strand (displaced within the R-loop complex).

  Despite the fact that type II CRISPR-Cas systems offer a strong defence against viral infection, phages can evade interference by employing a variety of evasion techniques. Effective targeting is prevented by spontaneous mutations in the PAM or protospacer89, and DNA modifications like glucosylation can reduce DNA recognition and/or increase mutation frequencies that allow escape. The discovery of numerous Cas9 phage-encoded inhibitors (also known as "anti-CRISPRs," discussed that halt nucleic acid-targeting at various stages by either preventing DNA binding or by impeding the conformational rearrangements necessary for DNA cleavage has also been made.

- **Cas-III System:** The targeting of nucleic acids in type III CRISPR-Cas systems, which involves three different nuclease modes, is likely the most challenging. Years of research suggested that the interference mechanism was inconsistent, with in vivo studies on a type III-A system reporting DNA targeting and in vitro studies on a type III-B system showing RNA targeting. This mystery was just recently answered, and it has since been expanded upon by astonishing findings on an additional nuclease activity that is triggered by the creation of second messengers. The multi-subunit effectors in type III systems are classified into four categories and are called Csm or Cmr complexes.

  As RNA-guided RNA-targeting effectors, type III complexes recognise single-stranded target RNAs by base-pairing with the crRNA. But a conserved -hairpin

found in Cas7 deliberately breaks up the resultant RNA duplex, causing a nucleotide to flip outward at 6-nt intervals. Thus, these particular nucleotides are set up for cleavage, which is made possible by a conserved aspartate residue found in Cas7.Rigid positioning of the crRNA handle within the effector complex orchestrates cleavage at fixed distances from the 5′ end, resulting in a characteristic, ruler-like degradation pattern. Type III immune systems have been demonstrated to target escape mutants from other CRISPR systems, in addition to utilising a formidable arsenal of weaponry for stand-alone defence. Additionally, they distinguish between lysogenic and lytic infections, enabling the host to benefit from possible advantages of lysogenic infection. These findings highlight the wonderful polishing influence of natural selection on the structure and operation of CRISPR-Cas systems.

- **Cas-IV System:** Type IV systems are a more recent addition. They express known multi-subunit protein-crRNA complex components but frequently lack adaptability genes, a potential DNA nuclease, and even CRISPR arrays. They have not been experimentally investigated, and it is unclear what they do.

- **Cas-V System:** Makarova et al. classified Type V systems as a possible grouping in 2015 based on the signature gene cpf1, which had been found in multiple bacterial genomes next to adaptation genes and a CRISPR array. Shortly after, Zetsche et al. published experiments that clearly established this new classification within Class 2 systems, showing that Cpf1—now known as Cas12a—works similarly to Cas9 as a single-effector, RNA-guided endonuclease that catalyses double-stranded DNA cleavage.

  Cas12(A-E) formerly known in the studies as Cpf1 (A), C2c1 (B), C2c3 (C), CasY (D), and CasX (E), are now part of Type V systems. All Cas12 homologs, with the exception of Cas12c, have been demonstrated to cause biochemical DNA cleavage and/or DNA interference activity in Escherichia coli. They all contain a RuvC nuclease domain but otherwise exhibit limited sequence similarity. The uncharacterized loci (subtype V-U) in the Type V group also encode Class 2 candidate proteins that are considerably smaller than other Cas12 family members. Given that they flank CRISPR arrays that frequently include spacers that match phage genomes, these putative RNA-guided effectors with RuvC-like nuclease domains are predicted to be active.

- **Cas-VI System:** Subtypes A-E of Type V systems, which encode Cas12a-e and were formerly known in the literature as Cpf1 (A), C2c1 (B), C2c3 (C), CasY (D), and CasX (E), have recently been added. Based on the discovery that they flank CRISPR arrays that frequently include spacers matching phage genomes, these putative RNA-guided effectors are projected to be active and exhibit RuvC-like

nuclease domains. Members of the Cas13 family possess two HEPN domains with expected single-stranded RNase activity, and tests using heterologously produced Type VI systems in E. coli showed that these systems offer protection against RNA phages. Upon binding to the target RNA, Cas13 activates, causing multiple-turnover nonspecific RNA cleavage activity to be released in trans rather than within the targeted region itself. Although not always, targeting is aided by the presence of a certain protospacer flanking sequence, and depending on the Cas13 homolog, collateral, nonspecific RNA cleavage is both sensitive to secondary structure and has particular nucleotide preferences.

The collateral, nonspecific cleavage of other cellular RNAs inhibits bacterial growth and may have developed as a method to cause cell dormancy and/or programmed cell death. Cas13 members may provide particular protection against RNA phages when produced heterologously in E. coli. Additional research is necessary to comprehend the physiological effects of interference in Type VI systems, especially involving assays that test Cas13 function in native species.

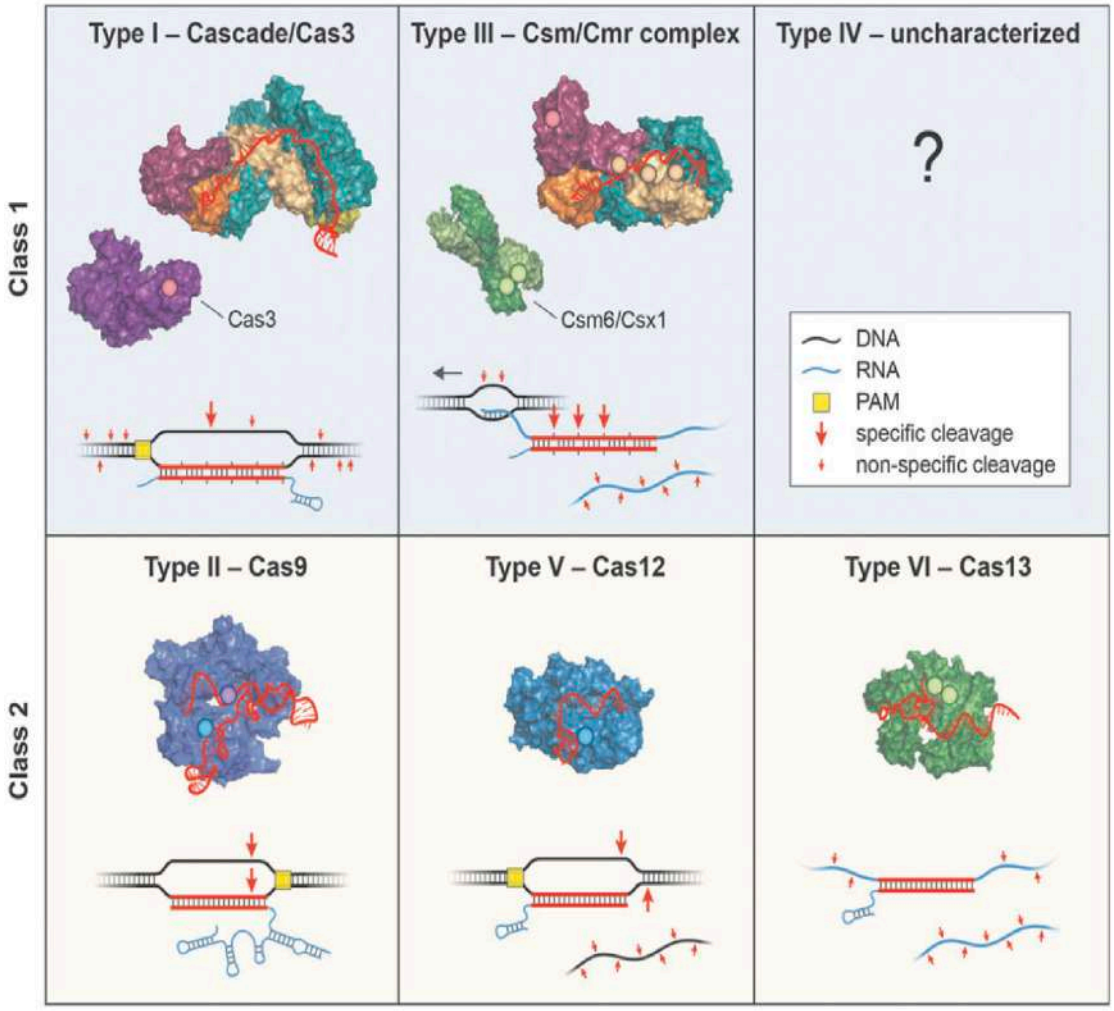


Fig: Different types of CRISPR-Cas system

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6636882/

Breaks repair mechanism: Cas9 may be designed to create DSBs at DNA target locations in conjunction with a natural dual-RNA guide or an artificial single-RNA guide49, leading to DNA repair outcomes that fall into two main classes: nonhomologous end joining (NHEJ) and homology-directed repair (HDR)185 (Fig. 3A). Despite not being random, repair results from NHEJ frequently lead to tiny insertions or deletions at the target site, which when introduced within exons frequently result in loss-of-function phenotypes. Combining Cas9-gRNA and a donor template enables access to precise genomic changes through HDR.

○ The potential impact of CRISPR Cas9 system

The CRISPR-Cas9 system is acclaimed as a revolutionary gene-editing tool in the modern scientific era. Among them, the most benefitted fields are-

Advancements in Treatment Procedures and Medical Research:

Medical research and treatment are improving due to the CRISPR-Cas9 system's capacity to accurately modify DNA sequences. It offers the possibility of specialised therapy for genetic illnesses that were previously thought to be incurable. It may contribute to revolutionary medicines for illnesses including cystic fibrosis, sickle cell anaemia, and particular types of cancer by addressing disease-causing mutations at their source. The CRISPR/Cas9 technology offers a powerful and innovative way to address conditions like diabetes. There are many approaches used, including cell-based therapies (such as stem cells and brown adipocytes), focusing on key genes involved in the aetiology of diabetes, and examining the difficulties and restrictions posed by this technology.

In a recent experiment, NBS-CRISPR, a sequence that interacts with NF-B, was recently tacked onto the CRISPR-Cas9 framework to create a system that causes inflammation. This newly created genome-editing tool's genetic scissor function is triggered in the presence of an inflammatory attack and deactivated or scaled back in non-inflammatory settings. Furthermore, by focusing on the MyD88 gene, a significant participant in the NF-B signalling system, they used this platform to correct inflammatory diseases and produced remarkable therapeutic results. Finally, P65 (RELA) can move from the cytoplasm to the nucleus during inflammation. They created an NBS-P65-CRISPR system expressing the Cas9-p65 fusion protein to prevent Cas9 leaky DNA cleavage activity. This inflammation-inducible Cas9-mediated

genome editing approach offers fresh perspectives and access points for investigating pathogenic genes.

→ Application and progress of Cas-9 based on hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is currently a common cause of cancer death despite no appreciable advances in patient survival over the previous few years. Despite the fact that surgery, chemotherapy, and radiotherapy have been the main traditional therapeutic modalities for HCC patients and have all made noteworthy progress, there are still issues, such as medication resistance and toxicity. The recently developed gene therapy of clustered regularly interspaced short palindromic repeat/CRISPR-associated nuclease 9-based (CRISPR/Cas9) has attracted considerable attention as an alternative to conventional treatment methods for eradicating resistant malignant tumours and controlling numerous crucial events of target gene-editing. At the intersection of several scientific fields, including chemistry, materials science, cancer biology, and genetics, recent discoveries in CRISPR/Cas9-based anti-drugs are explored. The CRISPR/Cas9 technique's basic principles and working examples were presented in some reviews to demonstrate its viability. The applications of CRISPR/Cas9 technology in therapeutic HCC were also highlighted. The design, action mechanisms, and anticancer uses of non-viral delivery systems for CRISPR/Cas9-based anti-drugs in HCC treatment were also summarised. Finally, the shortcomings and future directions of recent research were also examined in an effort to give a thorough theoretical foundation for the development of anti-drugs.

- **Customizing Drug Development**: The system's precision allows researchers to engineer cellular models with specific genetic traits, accelerating drug development. CRISPR-Cas9 facilitates the creation of more accurate disease models, enabling researchers to study disease mechanisms, test potential treatments, and evaluate drug safety more efficiently.

  A 2021 research used CRISPR-Cas9 and transcriptomic analysis to develop p53 knockout U87MG cell line for an unbiased drug delivery testing system. n their study, they used the CRISPR-Cas9 technology to generate various variants of U87MG cells with targeted mutations in the TP53 gene. They found that the loss of p53 function is the primary cause of their significant transcriptional changes. this information made them predict which mutant clones will have less divergent characteristics from the wild type and, thus, make the best candidates to be used as drug delivery testing platforms based on the transcriptome data. Their expectations were confirmed by additional in vitro and in vivo tests of cell shape, proliferative rate, and target antigen-mediated

uptake. The experiment was a success in choosing the top qualifying mutant clone based on the combined analysis results. This work establishes the methodology's viability and lays the door for expanding to new cell types and target genes.

- **Agriculture and Food Security**: The CRISPR-Cas9 system has the capability to develop crops with increased resistance to pests and diseases, enhanced nutritional value, and longer shelf life. This method can be helpful in addressing challenges with food security globally and may lead to lowering the impact of agriculture on the environment while reducing the need of chemical pesticides and insecticides.

  A recent study shows the progression of CRISPR/Cas9-based genome editing for improving disease resistance properties in plants. They are using strategies like transgenic technology, and traditional breeding. In recent years production of new lines with improved pathogen resistance was done. But there have been some drawbacks, including the transfer of undesirable genes along with desired resistance genes, the cost of labour, and the inherent genetic variability in plant populations. It's interesting that recent breeding techniques have made it possible to specifically modify one or more genes in plants. In the case of viral illnesses, the CRISPR/Cas9 technology was successfully exploited to increase plant biotic resistance via mutation of susceptibility (S) genes, detrimental regulators of the plant immune system, as well as pathogen genomes themselves.

- **Industrial and biotechnological applications**: The use of CRISPR-Cas9 has expanded the biotechnology field. It renders the ability to more effectively manipulate microbes to produce biofuels, medicines, and chemicals. This technology has the ability to foster environmentally friendly and sustainable solutions across numerous industries.

  Recent studies give an overview of the theories and most recent advances in CRISPR-Cas9 tools for multiplex genome editing, precise alteration of endogenous genes, and total knockout of their expression. The CRISPR-Cas9 system's present and future uses for cell imaging, metabolite engineering, biocontrol, chromatin dynamics, various signalling cascades, and deciphering fungal disease are also emphasised. Their contribution in discussing the difficulties encountered in the design and use of the CRISPR-Cas9 system is revolutionary in providing the problems of using these methods to study filamentous fungi.

- **Responsibilities and ethical considerations**: Although the previous explanation suggests CRISPR/Cas9 is a promising approach, the system also comes with some limitations and risks that make its use in clinical trials challenging. There are delivery challenges involved with CRISPR/Cas9. It is crucial to pick the optimum delivery method for introducing the CRISPR system into cells and to focus on the correct nucleus sequences, especially in vivo.

Important moral and societal issues are raised with the revolutionary potential of CRISPR-Cas9 system. Editing the human germline, for instance, raises concerns over the unexpected implications of genetic manipulation and its ethical limits. As this technology develops, it is crucial to ensure responsible use and take ethical issues into consideration.

Using Interference Complexes in the CRISPR-Cas9 system:

The majority of Cas protein tool development has been concentrated on making use of interference complexes' programmable, sequence-specific DNA recognition. In particular, Cas9 from S. pyogenes has shown to be incredibly helpful for genome engineering. Cas9 is now a two-component system that is simple to utilise for genome editing, transcriptional regulation, RNA targeting, and imaging due to having the ability to combine the crRNA and tracrRNA into a single guide RNA (sgRNA). From primary human T cells and stem cells to mice, primates, plants, bacteria, and fungus, Cas9 has been employed in a variety of cell types and creatures. Recent research has concentrated on using Cas9 orthologs with varied PAM sequences and smaller, chemical- and light-inducible Cas9 constructs to enable more spatiotemporal control. Other interference complexes can also be employed to modify the genome and have been used in the past.

Despite being less tractable for genome engineering than Cas9 due to its multi-subunit structure, Cascade has been employed for transcriptional silencing in E. coli due to its large size and stable binding. Although neither Csm nor Cmr complexes have been employed in published work, both may probably be used for a variety of cellular RNA manipulation applications. Out of the 16 Cpf1 homologs studied, two have been found to allow genome editing in human cells. The staggered cuts might turn out to favour different methods of DNA repair, and the alternative PAM specificity of Cpf1 might prove valuable for locating locations lacking an adequate PAM for Cas9.

The future of CRISPR-Cas9:

While Cas9 has already seen extensive use in the research setting, challenges remain for its application in the clinical fields. Even while doing programmed cuts has mostly become straightforward, it is still difficult to bias DNA repair in favour of homology-directed repair rather than non-homologous end joining. Delivery of Cas9, either as an RNP or on a plasmid or viral vector, to particular tissues in whole organisms is another challenge that must be addressed to enable clinical applications. Clinical studies might take place in a few years, and medicines might follow in a decade, as the field is still developing quickly. Although Cas9 has previously been extensively used in research settings, obstacles still stand in the way of its clinical application. Even while doing programmed cuts has mostly become straightforward, it is still difficult to bias DNA repair in favour of homology-directed repair rather than non-homologous end joining. Another difficulty that needs to be

resolved to allow for clinical applications is the delivery of Cas9 to specific tissues in complete organisms, whether it is via an RNP, a plasmid, a viral vector, or both. Clinical studies might take place in a few years, and medicines might follow in a decade, as the field is still developing quickly. Although Cas9 has previously been extensively used in research settings, obstacles still stand in the way of its clinical application. Even while doing programmed cuts has mostly become straightforward, it is still difficult to bias DNA repair in favour of homology-directed repair rather than non-homologous end joining. Another difficulty that needs to be resolved to allow for clinical applications is the delivery of Cas9 to specific tissues in complete organisms, whether it is via an RNP, a plasmid, a viral vector, or both. Clinical studies might take place in a few years, and medicines might follow in a decade, as the field is still developing quickly.

References:

1. Al-Attar, Sinan, Westra, Edze R., van der Oost, John and Brouns, Stan J.J.. "Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defence mechanism in prokaryotes" , vol. 392, no. 4, 2011, pp. 277-289.

2. Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., & Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA biology*, *16*(4), 504–517.

3. Boubakri, Hatem. "Recent Progress in CRISPR/Cas9-Based Genome Editing for Enhancing Plant Disease Resistance." *Gene* 866 (May 25, 2023): 147334.

4. "CRISPR and Pre-CRISPR RNA (CrRNAs) | Molecular Biology | JoVE." Accessed September 3, 2023

5. Nishimasu, Hiroshi, F. Ann Ran, Patrick D. Hsu, Silvana Konermann, Soraya I. Shehata, Naoshi Dohmae, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. "Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA." *Cell* 156, no. 5 (February 27, 2014): 935–49.

6. Guo, Minghui, Kuan Ren, Yuwei Zhu, Ziyun Tang, Yuhang Wang, Bailing Zhang, and Zhiwei Huang. "Structural Insights into a High Fidelity Variant of SpCas9." *Cell Research* 29 (January 21, 2019.

7. Babu, Kesavan, Nadia Amrani, Wei Jiang, S.D. Yogesha, Richard Nguyen, Peter Z. Qin, and Rakhi Rajan. "Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance." *Biochemistry* 58, no. 14 (April 9, 2019): 1905–17.

8. Babu, Kesavan, Nadia Amrani, Wei Jiang, S. D. Yogesha, Richard Nguyen, Peter Z. Qin, and Rakhi Rajan. "Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance." *Biochemistry* 58, no. 14 (April 9, 2019): 1905–17.

9. Anders, Carolin, Ole Niewoehner, Alessia Duerst, and Martin Jinek. "Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease." *Nature* 513, no. 7519 (September 25, 2014): 569–73.

10. Zuo, Zhicheng, and Jin Liu. "Structure and Dynamics of Cas9 HNH Domain Catalytic State." *Scientific Reports* 7 (December 8, 2017): 17271.

11. Glasgow, Anum, Helen T. Hobbs, Zion R. Perry, Malcolm L. Wells, Susan Marqusee, and Tanja Kortemme. "Ligand-Specific Changes in Conformational Flexibility Mediate Long-Range Allostery in the Lac Repressor." *Nature Communications* 14, no. 1 (March 2, 2023): 1179.
12. Synthego. "The Complete Guide to Understanding CRISPR SgRNA." Accessed September 3, 2023.

13. “Cas9.” In *Wikipedia*, August 25, 2023.

14. Yuan, Tingting, Honglin Tang, Xiaojie Xu, Jingjing Shao, Gaojun Wu, Young-Chang Cho, Yuan Ping, and Guang Liang. “Inflammation Conditional Genome Editing Mediated by the CRISPR-Cas9 System.” *IScience* 26, no. 6 (June 16, 2023): 106872.

15. Cheng, Yan, Haiyang Wang, and Mo Li. “The Promise of CRISPR/Cas9 Technology in Diabetes Mellitus Therapy: How Gene Editing Is Revolutionizing Diabetes Research and Treatment.” *Journal of Diabetes and Its Complications* 37, no. 8 (August 1, 2023): 108524.

16. Kwon, Andrew Tae-Jun, Kohta Mohri, Satoshi Takizawa, Takahiro Arakawa, Maiko Takahashi, Bogumil Kaczkowski, Masaaki Furuno, et al. “Development of P53 Knockout U87MG Cell Line for Unbiased Drug Delivery Testing System Using CRISPR-Cas9 and Transcriptomic Analysis.” *Journal of Biotechnology* 332 (May 20, 2021): 72–82.

17. Yu, Shijing, Ruirui Zhao, Bingchen Zhang, Chunmei Lai, Linyan Li, Jiangwen Shen, Xiarong Tan, and Jingwei Shao. “Research Progress and Application of the CRISPR/Cas9 Gene-Editing Technology Based on Hepatocellular Carcinoma.” *Asian Journal of Pharmaceutical Sciences* 18, no. 4 (July 1, 2023): 100828.

18. Klompe, Sanne E., and Samuel H. Sternberg. “Harnessing ‘A Billion Years of Experimentation’: The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems.” *The CRISPR Journal* 1, no. 2 (April 1, 2018): 141–58.