Department of BSBE
Indian Institute Of Technology Guwahati

Recent advances and applications: ZFNs, TALENs, CRISPR/Cas systems.

BT 207

Genetic Engineering

Jan-April 2023

CRISPR - CAS9

CRISPR-Clustered regularly interspaced short palindromic repeats

CAS9 –CRISPR associated protein 9

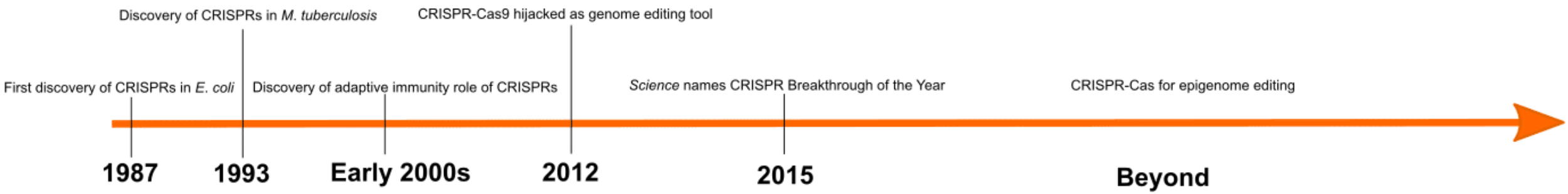
Zinc Finger Nuclease

TALEN - Transcription activator-like effector nuclease

Comparison of ZFN, TALEN and CRISPR/CAS

Properties	ZFN	TALEN	CRISPR/CAS
Abbreviation	Zinc Finger Nucleases	Transcription activator-like effector nucleases	Clustered, regularly interspaced, short palindromic repeat
Off target effect	Low	Low	High
Time Consumption	Long(7-15 days)	Long(5-7 days)	Short(1-3 days)
Cost	High	High	High
Essential Components	Zinc finger proteins + FokI fusion protein	TALE and FokI fusion protein	Guide RNA + Cas9 protein
Design Components	Protein	Protein	DNA
Cleaving strand	Double Stranded cleavage	Double Stranded cleavage	Single Stranded cleavage
Efficiency	Poor-33%	33%	High-70%
Methylation Sensitivity	Sensitive	Sensitive	No

The CRISPR timeline:
from discovery to genome editing and beyond

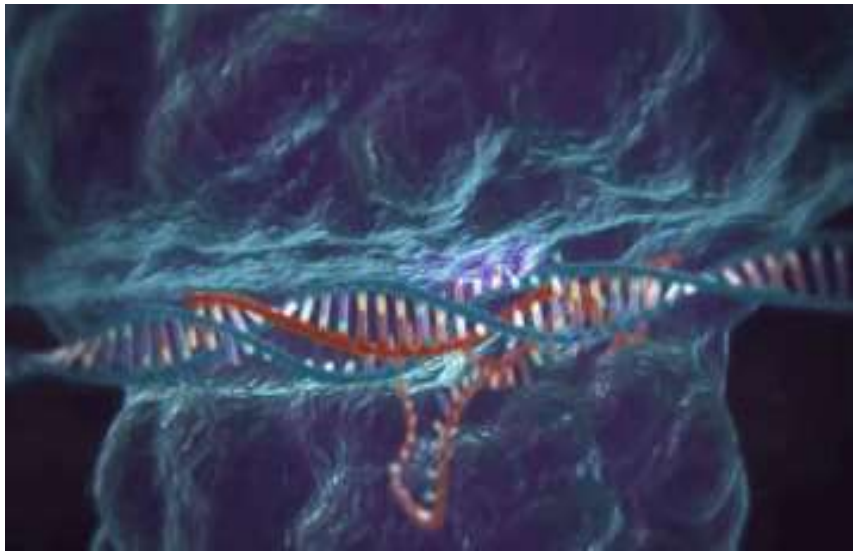


Key points in the history of CRISPR-Cas9:

- The CRISPR-Cas9 system was first discovered in 1987 by Japanese researcher Yoshizumi Ishino, who found repeating sequences in the DNA of the bacteria *Escherichia coli*.
- In 2005, researchers at the University of Alicante in Spain proposed that the CRISPR sequences might be involved in a defense mechanism against viral infection.
- In 2007, a group of scientists at Danisco, a food ingredients company in Denmark, showed that the CRISPR system could be used to protect bacteria from phage infection.
- In 2011, Jennifer Doudna at UC Berkeley and Emmanuelle Charpentier at Umeå University in Sweden collaborated on a project to understand the molecular mechanism of the CRISPR-Cas9 system.
- In 2012, Doudna and Charpentier published a paper in Science showing that the CRISPR-Cas9 system could be used to cut DNA in a sequence-specific manner.
- In the same year, Feng Zhang at the Broad Institute in Cambridge, MA, and George Church at Harvard University independently showed that the CRISPR-Cas9 system could be adapted for use in mammalian cells.
- Since then, the CRISPR-Cas9 system has been widely used for genome editing, and has revolutionized the field of molecular biology.
- In 2020, Doudna and Charpentier were awarded the Nobel Prize in Chemistry for their work on the CRISPR-Cas9 system

CRISPR/Cas

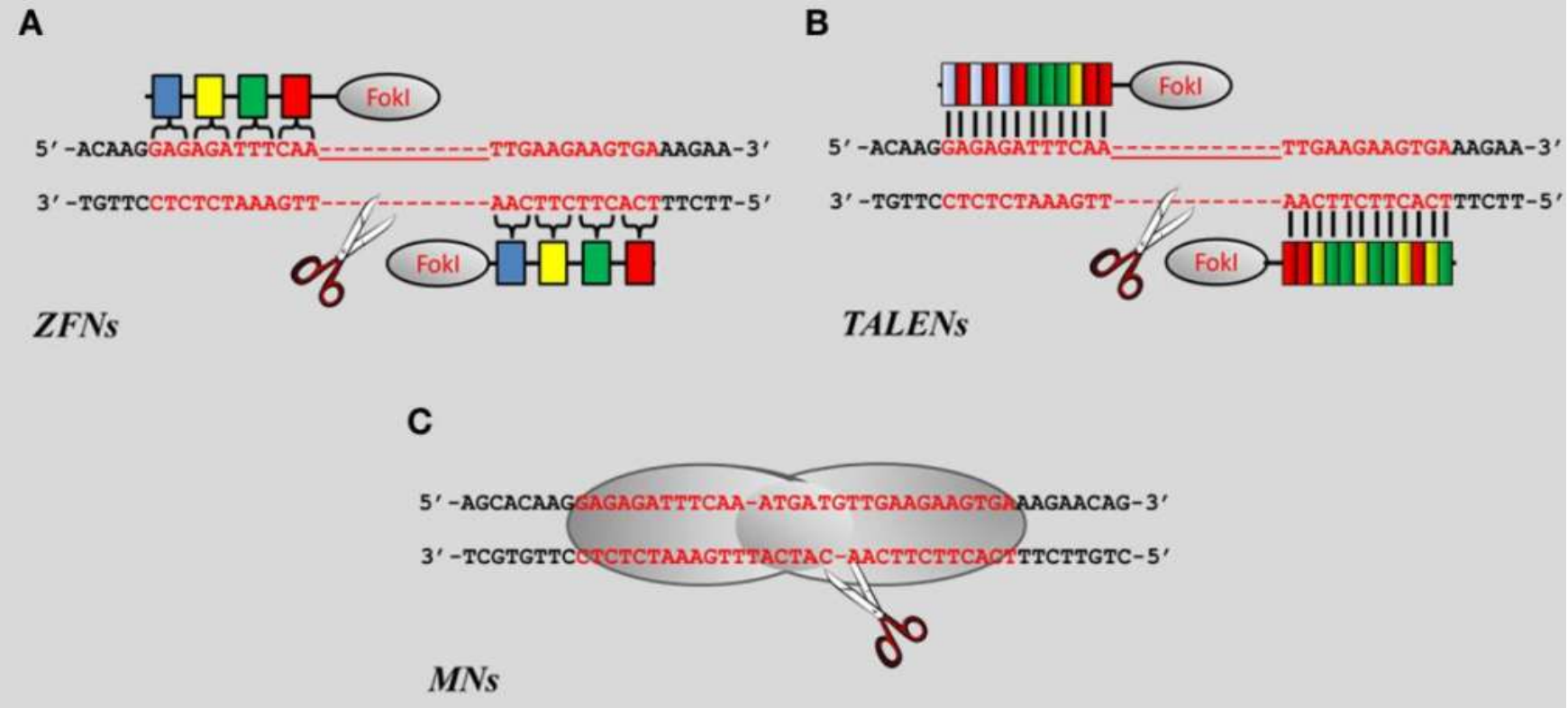
- **CRISPR- Cas9** : is clustered, regularly interspaced, short palindromic repeat and the associated Cas9 protein is naturally **an adaptive immunity mechanism** in prokaryotes.
- **CRISPRs** are found in approximately **40%** of sequenced **bacterial** genomes and **90%** of sequenced **archaea**.
- **Cas 9** is an **endonuclease**, first identified from ***Streptococcus pyogenes*** bacteria. Its **genes** are often **located next to CRISPR** repeat-spacer arrays.
- Specificity of **CRISPR-Cas9** depends on the presence of a sequence-specific Protospacer Adjacent Motif (**PAM**) and **target sequence** (20 bases).
- **Absence** of **PAM** in **host genome** enable to avoid **self-cleavage.**



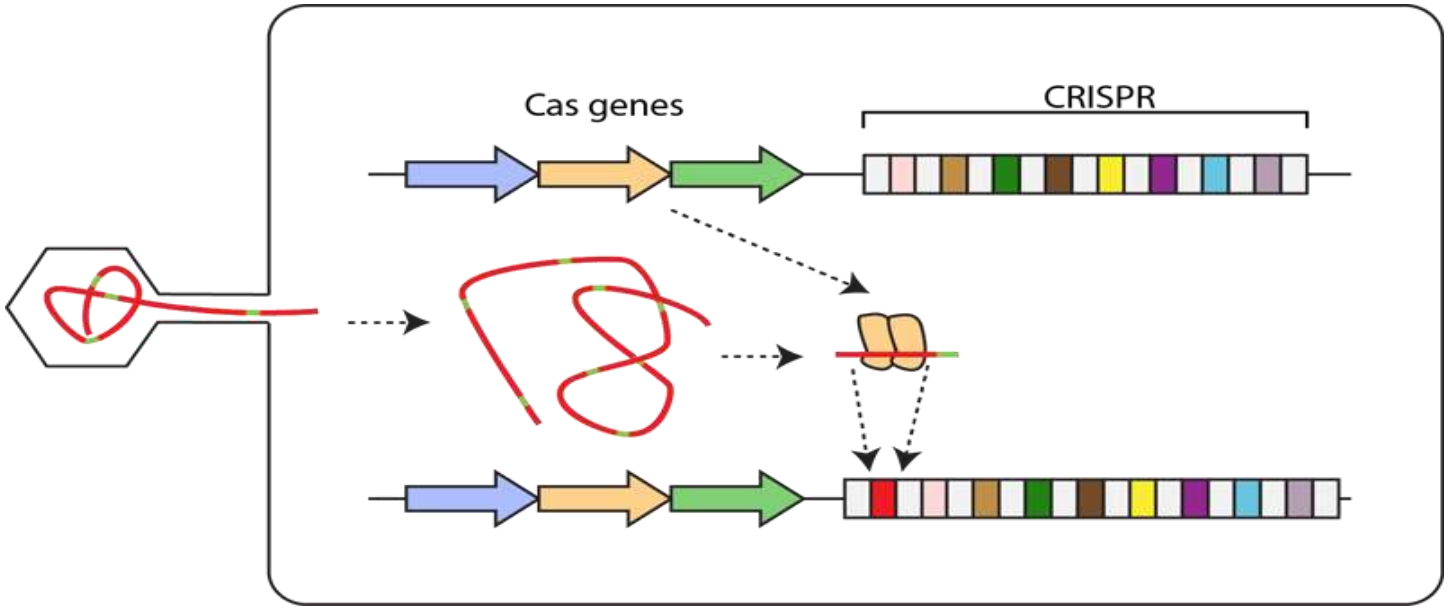
- **CRISPR** - an acronym for **clustered regularly interspaced short palindromic repeats** is a family of DNA sequences found in the genomes of prokaryotic organisms such as bacteria and archaea.
- These sequences are derived from DNA fragments of bacteriophages that had previously infected the prokaryote.
- They are used to detect and destroy DNA from similar bacteriophages during subsequent infections. Hence these sequences play a key role in the antiviral (i.e. anti-phage) defense system of prokaryotes and provide a form of acquired immunity.
- CRISPR is found in approximately 50% of sequenced bacterial genomes and nearly 90% of sequenced archaea.
- Cas9 (or "CRISPR-associated protein 9") is an enzyme that uses CRISPR sequences as a guide to recognize and cleave specific strands of DNA that are complementary to the CRISPR sequence.
- Cas9 enzymes together with CRISPR sequences form the basis of a technology known as CRISPR-Cas9 that can be used to edit genes within organisms.
- This editing process has a wide variety of applications including basic biological research, development of biotechnological products, and treatment of diseases

Introduction

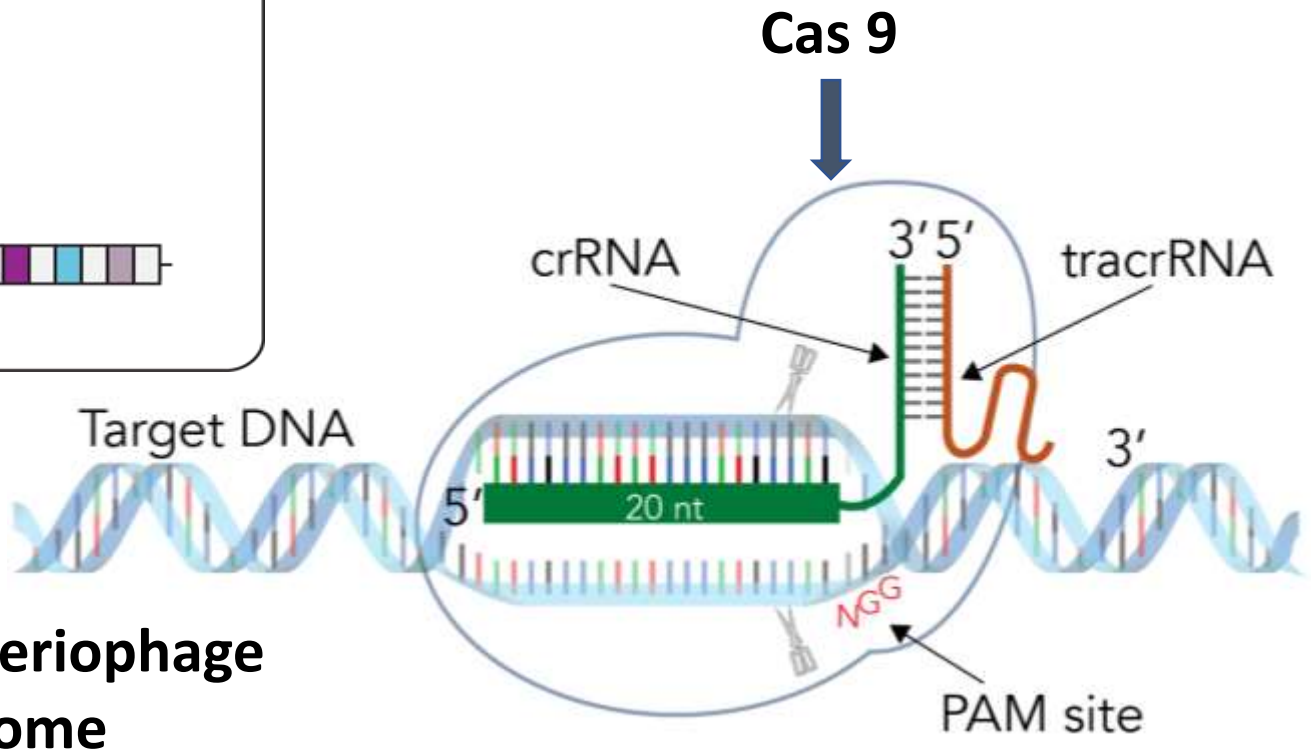
- Unlike **gene editing tools** like TALEN, Zink finger and Meganuclease which needs heavy protein engineering,
- **CRISPR- Cas9** is based on **programmable RNA guide** system. Its **easier** to apply and **inexpensive**.



Components of CRISPER-Cas9 System



Host Genome



Bacteriophage Genome

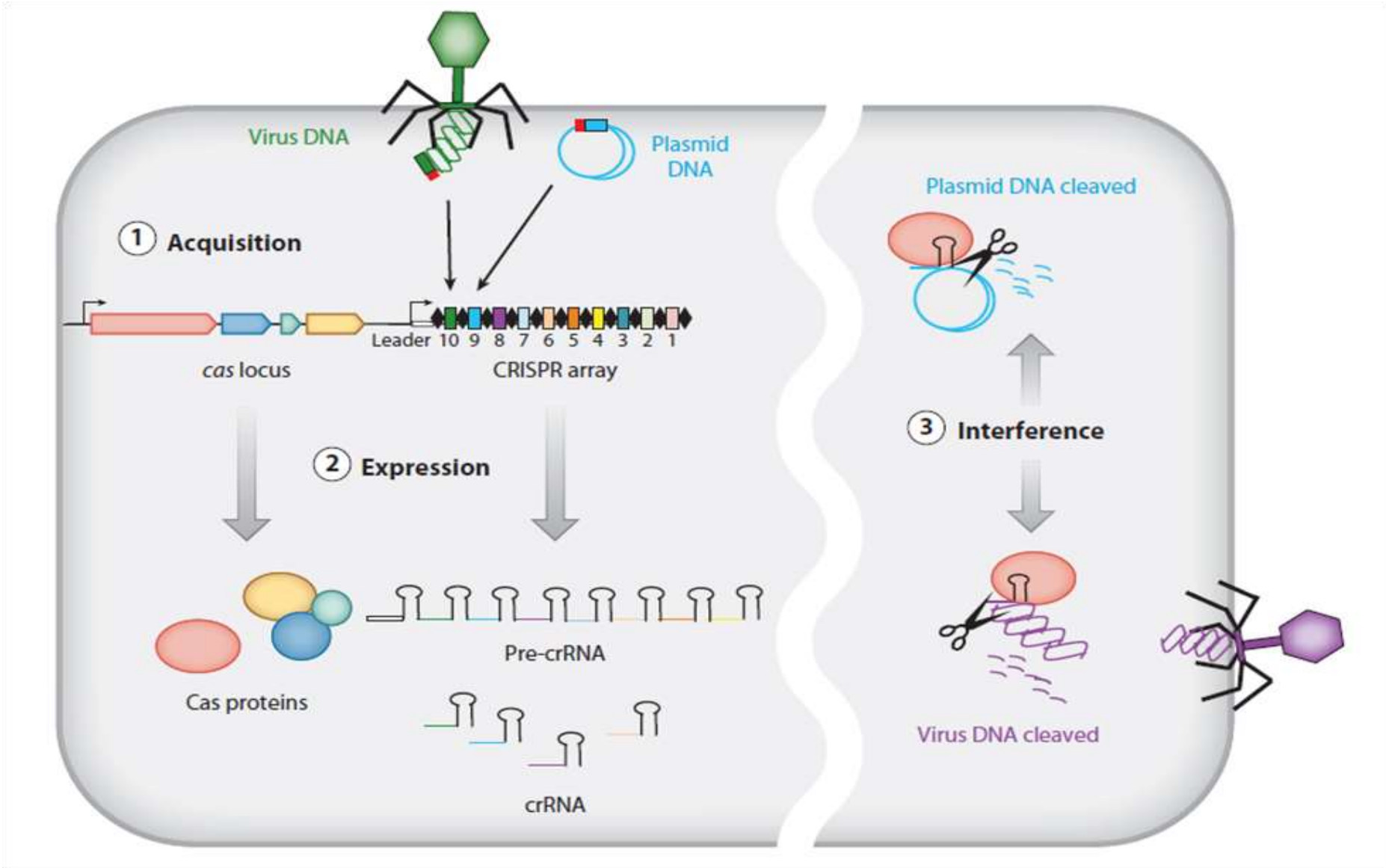
Three Stages of CRISPR-Cas9

- **Stage I: Acquisition- DNA fragments** of invading phages are **incorporated** into the host **CRISPR locus as spacers between crRNA repeats**.

- **Stage II: Expression/Biogenesis- Cas proteins** are **expressed**, the CRISPR array containing **acquired spacers** is **transcribed** into **pre-crRNA**, and then **cleaved** into **mature crRNAs**.

- **Stage III: Cleavage/Interference- Cas proteins recognize** the appropriate **target** with the **guidance** of the **crRNA** and mediate the **cleavage** of the invading genome to **generate DSB** at **3 bp upstream** of **PAM** sequence.

The Principle Behind…



CRISPR–Cas system

Zinc Finger Nuclease

- Zinc finger nuclease (ZFN) is a class of artificially-engineered nuclease that consists of a series of zinc fingerprotein domains fused with the cleavage domain of nuclease Fok I (Fig. 1).

- The sequence specificity and binding affinity of ZFNs to the target DNA are determined by zinc finger protein domains derived from naturally-occurring eukaryotic zinc finger-containing proteins (ZFPs) that exhibit an unique ability of recognizing and binding to specific DNA sequences.

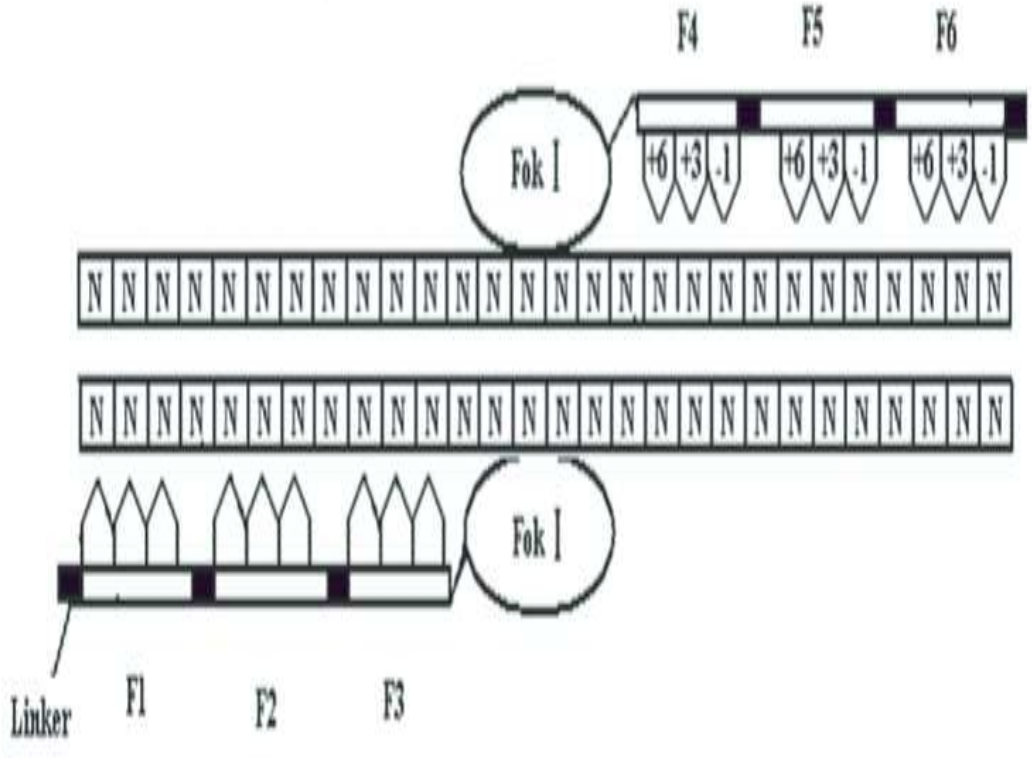


Fig 1. Structure of Zinc finger nuclease. Each monomer is composed of ZFA and Fok I nuclease, it generally works when two monomers form a dimer, F1–F6 are zinc fingers, each zinc finger is connected through the linker. +6,+3, and−1aminoacid of each zinc finger bind DNA.

- The series of zinc finger protein domains is also termed zinc finger motifs or zinc finger arrays.
- Based on the conserved amino acids in the zinc finger domain, ZFPs and ZFNs can be categorized into three major subtypes (C2H2,C4,andC6).
- Among these,C2H2is the classic type and the most widely used in the artificially-engineered ZFNs. TFIIIA was the first identified C2H2ZFP that consists of with zinc finger domain of 30 amino acids, in which the 8th and 13th amino acids of cysteine and the 26th and 30th amino acids of histidine are highly conserved throughout evolution.
- These conserved amino acids are responsible for the direct binding to zinc ion, which critical in stabilizing the protein fold for DNA binding. This type of Cys2His2 (C2H2)ZFP typically binds to DNA in the form of monomer.
- By comparison, the type of C4ZFP has 4 cysteines in as the con-served amino acid, which binds to DNA in a form of dimer. Reversed repeats in target DNA sequence are recognized by homodimer, whereas direct repeats are bound by heterodimer.
- The third type of C6ZFP has the ability to bind to 2 zinc ions by six conserved cysteines, which are critical in binding to DNA either in the form of monomer or dimer.

Application and recent advances:

- ZFN is the first engineered endonuclease to be used for gene editing.
- In the past few years, ZFNs have been widely used in gene editing of specific targets for gene knock-out, gene integration, and gene replacement.
- The latter two can be broadly considered as gene knock-ins. As one of the pioneer application, the ‘yellow’ gene was successfully knock-outed in Drosophila by ZFN via NHEJ.
- There are an ever-increasing number of model organisms in which genes are edited by ZFNs or other similar types of engineered endonucleases. Some of the species including Drosophila, zebrafish, C. elegans, mouse, rat, and pig.
- The major applications of gene editing using ZFNs can be divided into 3 major categories:

1. Disease modelling at cellular level or intact organismal level.
2. Genome modification of economically interested species for agriculture and industry
3. Human gene therapy.

TALEN - Transcription activator-like effector nuclease

- Transcription activator-like effector nuclease technology leverages artificial restriction enzymes generated by fusing a TAL effector DNA-binding domain to a DNA cleavage domain.
- Restriction enzymes are enzymes that cut DNA strands at a specific sequence. Transcription activator-like effectors (TALEs) can be engineered to bind practically any desired DNA sequence. By combining such an engineered TALE with a DNA cleavage domain (which cuts DNA strands), one can engineer restriction enzymes that will specifically cut any desired DNA sequence.
- When these restriction enzymes are introduced into cells, they can be used for gene editing or for genome editing in situ, a technique known as genome editing with engineered nucleases.
- Alongside zinc finger nucleases and Cas9 proteins, TALEN is becoming a prominent tool in the field of genome editing.

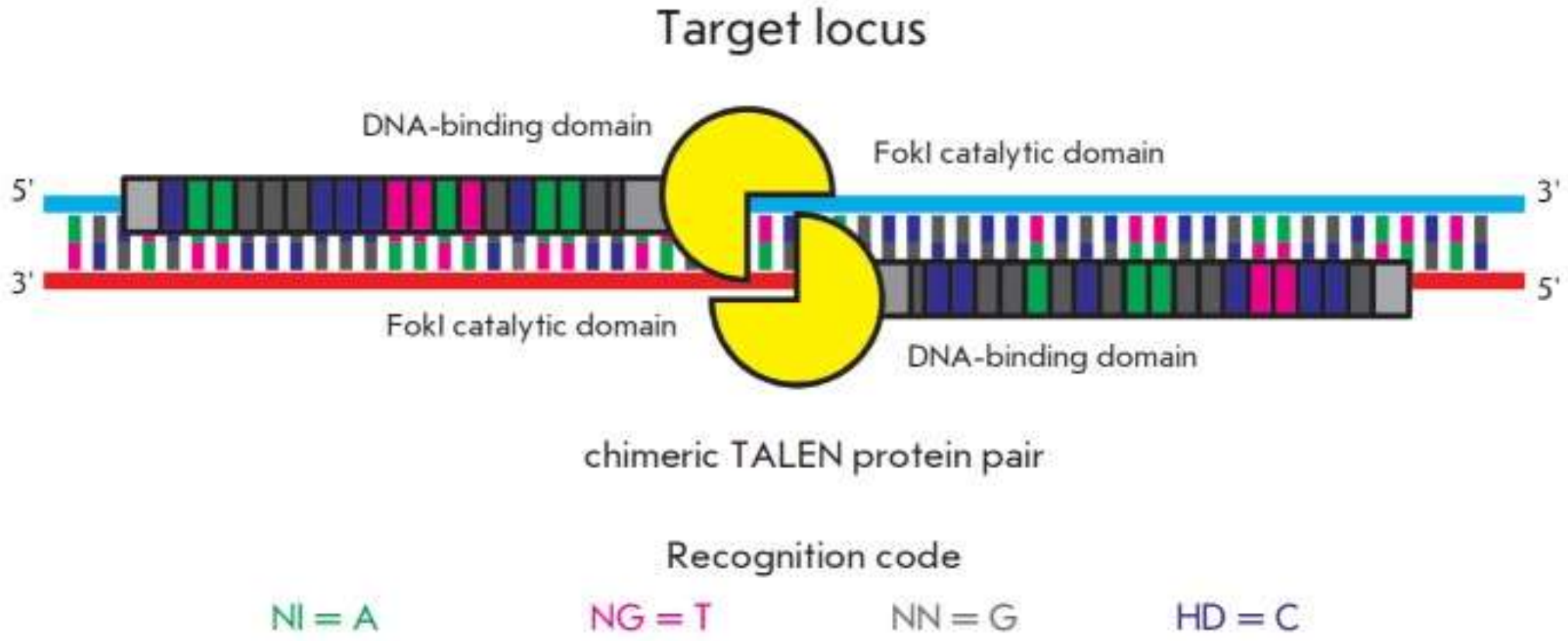


Fig. A scheme for introducing a double-strand break using chimeric TALEN proteins. One monomer of the DNA-binding protein domain recognizes one nucleotide of a target DNA sequence. Two amino acid residues in the monomer are responsible for binding. The recognition code (single-letter notation is used to designate amino acid residues) is provided. Recognition sites are located on the opposite DNA strands at a distance sufficient for dimerization of the FokI catalytic domains. Dimerized FokI introduces a double-strand break into DNA.

TAL effector DNA-binding domain

- TAL effectors are proteins that are secreted by *Xanthomonas* bacteria.
- The DNA binding domain contains a repeated highly conserved 33–34 amino acid sequence with divergent 12th and 13th amino acids.
- These two positions, referred to as the Repeat Variable Diresidue (RVD), are highly variable and show a strong correlation with specific nucleotide recognition.
- This relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA-binding domains by selecting a combination of repeat segments containing the appropriate RVDs.

DNA cleavage domain

- The non-specific DNA cleavage domain from the end of the FokI endonuclease can be used to construct hybrid nucleases that are active in many different cell types.
- The FokI domain functions as a dimer, requiring two constructs with unique DNA binding domains for sites in the target genome with proper orientation and spacing.
- Both the number of amino acid residues between the TALE DNA binding domain and the FokI cleavage domain and the number of bases between the two individual TALEN binding sites appear to be important parameters for achieving high levels of activity.

TALEN mechanism

- The simple relationship between amino acid sequence and DNA recognition of the TALE binding domain allows for the efficient engineering of proteins. Once the TALEN constructs have been assembled, they are inserted into plasmids; the target cells are then transfected with the plasmids, and the gene products are expressed and enter the nucleus to access the genome.

- Alternatively, TALEN constructs can be delivered to the cells as mRNAs, which removes the possibility of genomic integration of the TALEN-expressing protein. Using an mRNA vector can also dramatically increase the level of homology directed repair (HDR) and the success of introgression during gene editing.

- TALEN technology can be used to edit genomes by inducing double-strand breaks (DSB), which cells respond to with repair mechanisms.

- Non-homologous end joining (NHEJ) reconnects DNA from either side of a double-strand break where there is very little or no sequence overlap for annealing.

- This repair mechanism induces errors in the genome via insertion or deletion, or chromosomal rearrangement; any such errors may render the gene products coded at that location non-functional.

- Because this activity can vary depending on the species, cell type, target gene, and nuclease used, it should be monitored when designing new systems.

- Alternatively, DNA can be introduced into a genome through NHEJ in the presence of exogenous double-stranded DNA fragments. Homology directed repair can also introduce foreign DNA at the DSB as the transfected double-stranded sequences are used as templates for the repair enzymes.

Application and recent advances:

- TALEN technology has been used for instance to efficiently engineer stably modified human embryonic stem cell and induced pluripotent stem cell (IPSCs) clones and human erythroid cell lines.

- The technology has also been utilized experimentally to correct the genetic errors that underlie disease. For example, it has been used in vitro to correct the genetic defects that cause disorders such as sickle cell disease, xeroderma pigmentosum, and epidermolysis bullosa.

- It was also shown that TALEN technology can be used as tools to harness the immune system to fight cancers. In theory, the genome-wide specificity of engineered TALEN fusions allows for correction of errors at individual genetic loci via homology-directed repair from a correct exogenous template.

- In reality, however, the in situ application of TALEN technology is currently limited by the lack of an efficient delivery mechanism, unknown immunogenic factors, and uncertainty in the specificity of TALEN binding. Another emerging application of TALEN technology is its ability to combine with other genome engineering tools, such as meganucleases.

Questions