

Genome/Gene Editing

Subjects: [Genetics & Heredity](#)

Contributor: Wei-Ming Lin

Theoretically, a DNA sequence-specific recognition protein that can distinguish a DNA sequence equal to or more than 16 bp could be unique to mammalian genomes. Long-sequence-specific nucleases, such as naturally occurring Homing Endonucleases and artificially engineered ZFN, TALEN, and Cas9-sgRNA, have been developed and widely applied in genome editing.

CRISPR

Cas9

genome/gene-editing

extracellular vesicles

1. Preface

Besides the common gene knockout purpose, a major goal of genome/gene editing experiments is to precisely convert a selected DNA sequence into a new desired one in the native context of the whole genome. Before the advent of programmable nucleases, most of the cases of the site-specific changes by homology-directed repair (HDR) were executed under the single agent gene editing strategy utilizing single-stranded DNA oligonucleotide templates [1]. The efficiency of HDR could be highly improved when DNA cleavage occurred at or near the recombination site [2][3]. As a comparison of the ranges of genome sizes, those for amphibians and flowering plants are from 1E9 to 1E11 and from 1E8 to 1E11 bp, those of birds and mammals are quite narrow, being around 1E9 and 3E9 bp per ploidy, respectively [4][5]. If we want to precisely operate the genomes of livestock, such as cows, pigs, sheep, and chickens, a nuclease which is able to recognize DNA sequences of equal to or more than 16 bp, theoretically, is indispensable. (let $4^n \geq 3E9$, then $n \geq 16$). Natural or engineered site-specific nucleases suitable for these purposes and their applications on genome/gene editing are described in this paper.

2. Endonucleases for Genome Editing

2.1. Homing Endonuclease and Meganuclease

Since the original study in 1971 to determine the “ ω ” self-splicing intervening sequence, it was later recognized as a group I intron [6] located within the mitochondrial gene-encoding large ribosomal RNA of yeast, *Saccharomyces cerevisiae* [7]. Additionally, the inheritance of this intron was induced by a site-specific endonuclease (now termed I-SceI) encoded within the intron sequence [8]. Such nucleases which can recognize a unique DNA target site among the whole genome are now nomenclatured as Homing Endonucleases, defined as microbial DNA-cleaving enzymes that mobilize their own reading frames by generating double-strand breaks at specific genomic invasion sites. These proteins display an economy of size and yet recognize long DNA sequences (typically 20 to 30 base pairs) [9]. Based on the specific amino acid motif in the catalytic core, HEs were categorized into at least six

families, with each being associated with a particular host range. HNH, GIY-YIG, and EDxHD families are largely constrained to phage hosts. PD-(D/E)xK, His-Cys Box, and LAGLIDADG families are encoded in bacterial, protistic, and archaeal/eukaryotic hosts, respectively [10][11][12]. The endonucleases of the LAGLIDADG HE family are often referred to as “meganucleases (MN)” [13]. The LAGLIDADG endonucleases exist both as homodimers (where the two identical protein subunits are each typically 160 to 200 residues in size with a $\alpha\beta\beta\alpha\beta\beta\alpha$ core fold in which a long anti-parallel pair of β strands are fitted into the major groove of recognition site), and as monomeric proteins, where a tandem repeat of two LAGLIDADG domains is connected by a variable peptide linker. Compared with their homodimeric cousins, the monomeric proteins are rather small; their individual domains are often only 100 to 120 residues in size and can recognize fully asymmetric DNA target sites. In addition to the benefits that the recognition site of MNs are quite long (18–24 bp) and highly specific and the sizes are small so as to be prepared and engineered [14][15] easily, the richest natural sources are another advantage. For the full spectrum of the mammalian genome, a bank of 3 billion MNs is theoretically needed to cover all possible recognition sites, or hundreds of thousands to overlay all genes. That is a tremendous task.

2.2. Zinc Finger Nuclease (ZFN)

In comparison to a typical type II restriction enzyme, whose recognition site is overlapped with a cutting site, in the type IIS, where “S” means shifted cleavage, the enzyme contains independent modules for a separated recognition site and cutting site, e.g., FokI contains a 382 a.a. N-terminal DNA recognition domain and a 196 a.a. C-terminal nuclease domain (FN) [16]. After the proof-of-concept work prepared chimeric restriction endonuclease by linking the site-specific DNA-binding homeodomain of Ubx with FN [17], the primitive concept of combining the site-specific DNA-binding zinc finger domains with FN, termed as ZFN hereafter, to be an editable hybrid restriction enzyme was launched by Dr. Srinivasan Chandrasegaran’s group at Johns Hopkins [18]. The zinc finger is a big superfamily of domains and C_2H_2 is the most common type of it. A β sheet-turn- β sheet-turn- α helix has a rigid structure, where two cysteine residues located within the first turn and two histidine residues at the C-terminus of the α helix are coordinated to chelate a zinc ion. The α helix is fitted into the major groove of the DNA double helix and the -1, 3, and 6 a.a. residues interact with three consecutive nucleotides of the sense strand in the 3’ to 5’ orientation, respectively [19]. The C_2H_2 zinc fingers can be recognized as independent trinucleotide binding modules and linked into a polypeptide to distinguish longer DNA binding. In theory, one can design a zinc finger for each of the 64 possible combinations of trinucleotides, and one can arrange such zinc fingers to compose a sequence-specific artificial protein for any segment of DNA [20]. Because the nuclease activity of FN is stringently present in a dimer form [21], a pair of ZFNs with recognition sites in a tail-to-tail orientation was demonstrated necessary to perform effective double-strand cutting activity, which was essential to enhance the probability of homologous recombination a thousand-fold in vivo [22]. The D_{483} and R_{487} residues of FokI were involved in the FN dimer formation by interacting with each other between the two subunits. FN carried a $D_{483}R$ mutation which led both of the 483 and 487 residues of FokI to be positively charged, termed as FN_{RR} . On the other hand, FN carried a $R_{487}D$ mutation, which led both the 483 and 487 residues of FokI to be negatively charged, termed as FN_{DD} . Unlike the wild-type FN, FN_{RR} and FN_{DD} cannot form homodimers themselves; however, they can form heterodimers efficiently [23]. Such phenomena were utilized to improve the cutting specificity by using a pair of complementary FNs for the up-stream and down-stream recognition arm, respectively, as well as to reduce the toxicity caused by

the off-targeting by unwanted homo-dimer [24]. Although ZFN seems a powerful tool for genome editions, some drawbacks should be noted. Firstly, there are still no perfect matches between zinc finger proteins and DNA triplexes. The specificity between GC-rich triplexes and zinc fingers was calculated to be 73%, whereas it was only 50% for AT-rich triplexes and their zinc finger partners [25]. The DNA-binding specificity of a C₂H₂ zinc finger was also revealed to be influenced by neighboring ones [26].

2.3. Transcription Activator-Like Effector Nuclease (TALEN)

Since the isolation of the *avrBs3* gene in the bacterial plant pathogen *Xanthomonas campestris* pv. *Versicatoria* [27], AvrBs3 protein was found to be injected into plant cells via a type III secretion system [28][29] and to act as a site-specific transcription activator-like effector (TALE) to reprogram host cells. The AvrBs3 protein is composed of 1163 a.a. with a translocation domain at the most N-terminal end of a 287 a.a. N-terminal region, a middle part which contains 17 units of 34 a.a. complete repeats following with a 20 a.a. half repeat, and a 278 a.a. C-terminal region containing nuclear localization signals (NLSs) and an acidic transcriptional activation domain (AD) [30]. The sequences of the 34 a.a. repeats performing an α helix-random coil- α helix structure are nearly identical, except the polymorphic 12th and 13th residues, which are known as the repeat variable di-residue (RVD) and specifically specify a single binding site nucleotide through direct interactions [31][32][33][34]. The specificity and affinity for each RVD to a nucleotide were systemically studied [35][36]. Theoretically, merely four kinds of repeat, each with RVDs recognizing G, A, T, and C, respectively, are necessary to construct a DNA-binding domain specific to any given sequence. The N-terminus 152 a.a. and C-terminus 215 a.a. of AvrBs3 protein could be removed to leave a core region with intrinsic DNA-binding activity, and this core region was used as a fundamental framework for TALE effector nuclease (TALEN) designation [37][38][39][40]. X-ray crystal data revealed that the amino acid residues 162 to 288 of AvrBs4 perform four cryptic repeats of helical bundles to interact with a nucleotide T on the sense strand of a DNA double helix immediately in front of the nucleotides recognized by the canonical repeats, and this region provides the majority of the energy required for high-affinity target binding [41]. The RVDs of complete repeat modules bind consecutive nucleotides of sense strands in the 5' to 3' direction. The first residue in each RVD (the 12th of the repeat) orients away from DNA to interact with the backbone of the eighth residue of the repeat to stabilize the interhelical loop and allow the second residue of the RVD to project into the major groove of the DNA and make sequence-specific contact with a single nucleotide of the sense strand [42]. The most common RVDs are HD, NG, NN, and NI for C, T, G > A, and A, respectively, in which NN can be replaced by NH for NH is more specific to G but has less affinity [34][35][36][43]. Online tools for custom TALEs and TALENs, such as TALE-NT 2.0, designation and modular assembly methods relying on Golden Gate cloning have been developed, enabling researchers to make constructs in a few days [44][45][46][47][48]. Besides the FN nuclease domain, transcription regulatory domains and DNA modification enzymes can be engineered to the C-terminus of the sequence-specific TALE-core structure to create artificial gene regulatory factors [49][50].

2.4. CRISPR/Cas Nucleases

It was identified that the spacer sequences between identical repeats of the clustered regularly interspaced short palindromic repeat (CRISPR) loci of bacterial genomes might originate from plasmid and phage [51][52]. The CRISPR RNA and CRISPR-associated protein (Cas) systems are now confessed as key components governing

bacterial adaptive immune response which consists of three main stages: adaptation, expression, and interference. When a bacterium was attacked by an invader, a short DNA fragment, termed a protospacer, which is neighbored by a protospacer-adjacent motif (PAM) of the invader, was processed by adaptation Cas members, such as Cas1 and Cas2, to be inserted into the 5' end of a spacer-repeat CRISPR array embedded in the host genome as stored memory. Memory was retrieved as the CRISPR array was transcribed to produce a long precursor CRISPR RNA (pre-crRNA), which was processed by an expression factor, such as Cas6 or RNase III, within the repeat region to create mature crRNA, which was incorporated with Cas effectors, such as Cas5, Cas7, Cas8, and Cas11, to yield an RNA-guided sequence-specific endonuclease in the interference stage [53][54]. According to the number of Cas protein subunits included in the effector endonuclease complex, the CRISPR-Cas systems belong to two classes, with multi-subunit effector complexes in class 1, which can be further divided into three types: type I, type III and type IV, and single-protein effectors in class 2, including type II, type V and type VI [55][56][57]. Besides crRNA and Cas9 protein, a trans-activating CRISPR RNA (tracrRNA) whose 5' region is complemented with the repeat sequence of crRNA is critical to perform endonuclease activity in the type II CRISPR systems. The crRNA and tracrRNA could be engineered into one single-guided RNA (sgRNA) in accompaniment with Cas9 to restore full and specific endonuclease activity [58]. The best-characterized and applied Cas9 enzyme was originally isolated from *Streptococcus pyogenes*, and was referred to as SpCas9, or even simply as Cas9. SpCas9 is a large 1368 a.a. multidomain protein with two distinct lobes: the recognition (REC) lobe and the nuclease (NUC) lobe, connected through an arginine-rich bridge helix (residue 56 to 93) and a disordered loop (residue 712 to 717). The REC lobe is composed of three α -helical domains (Hel-I, Hel-II, and Hel-III) and the NUC lobe contains HNH and RuvC-like nuclease domains, as well as a PAM-interacting (PI) C-terminal domain [59][60] (**Figure 1A,B**). The apo-Cas9 protein should be assembled with guide RNA (native crRNA-tracrRNA hybrid or sgRNA) to achieve site-specific DNA recognition and cleavage activities. The 20 nt spacer sequence of crRNA provided DNA target specificity and the tracrRNA conferred a crucial role in Cas9 protein recruitment. Once the PAM (NGG for SpCas9) directly adjacent to a protospacer target site was trapped by R1333 and R1335 of the Cas9-guide RNA complex, it triggered local DNA melting at the PAM-adjacent site. The PAM-proximal 10–12 nucleotides (nt), 3'-end of the 20 nt spacer sequence is absolutely critical for site specificity, and was referred to as seed region. The DNA cleavage activity of CRISPR-Cas9 was excited by the conformational change induced by the R-loop formation between target DNA and spacer RNA [61][62]. The target DNA strand complementary to spacer RNA was cut by the HNH nuclease domain and the non-target DNA strand by the RuvC nuclease domain to produce a blunt-ended double-strand breakage at 3 bp upstream to PAM [63][64]. Either D₁₀A [58] or H₉₈₃A [65] mutation destroyed the RuvC nuclease activity. On the other hand, D₈₃₉A [66], H₈₄₀A [58], and N₈₆₃A [67] mutations could eliminate the HNH nuclease activity. These mutations did not influence the target site binding affinity of Cas9-sgRNA. Cas9 carrying the D₁₀A mutation and D₁₀A/H₈₄₀A double mutations were termed nickase (nCas9) and dead enzyme (dCas9), respectively (**Table 1**). The dCas9 could be taken as a guide RNA-derived sequence-specific DNA-binding protein, like TALE described above, and coupled with DNA manipulation enzymes or transcriptional activating/inhibitory domains to be harnessed for various applications [64]. The amino acid residues interacting with the PAM bases could be engineered to generate new PAM so as to broaden the spectrum of target sites. Based on the structure-guided rational design, the wild-type D1135, R1335, and T1337 were converted to E, Q, and R, respectively; the PAM was shifted from NGG to NGA. Additionally, as D1135, G1218, R1335, and T1337 were converted to V, R, E,

and R, respectively; the PAM became NGC [68]. An engineered SpCas9 bearing D₁₁₃₅L/S₁₁₃₆W/G₁₂₁₈Q/E₁₂₁₉Q/R₁₃₃₅Q/T₁₃₃₇R substitutions in PI domain (SpG) targeted NGN PAM. SpG was further engineered to carry A₆₁R/L₁₁₁₁R/N₁₃₁₇R/A₁₃₂₂R/R₁₃₃₃P substitutions to near-PAMless (NRN > NYN) variants, termed SpRY, with full endonuclease activities [69] (Table 1).

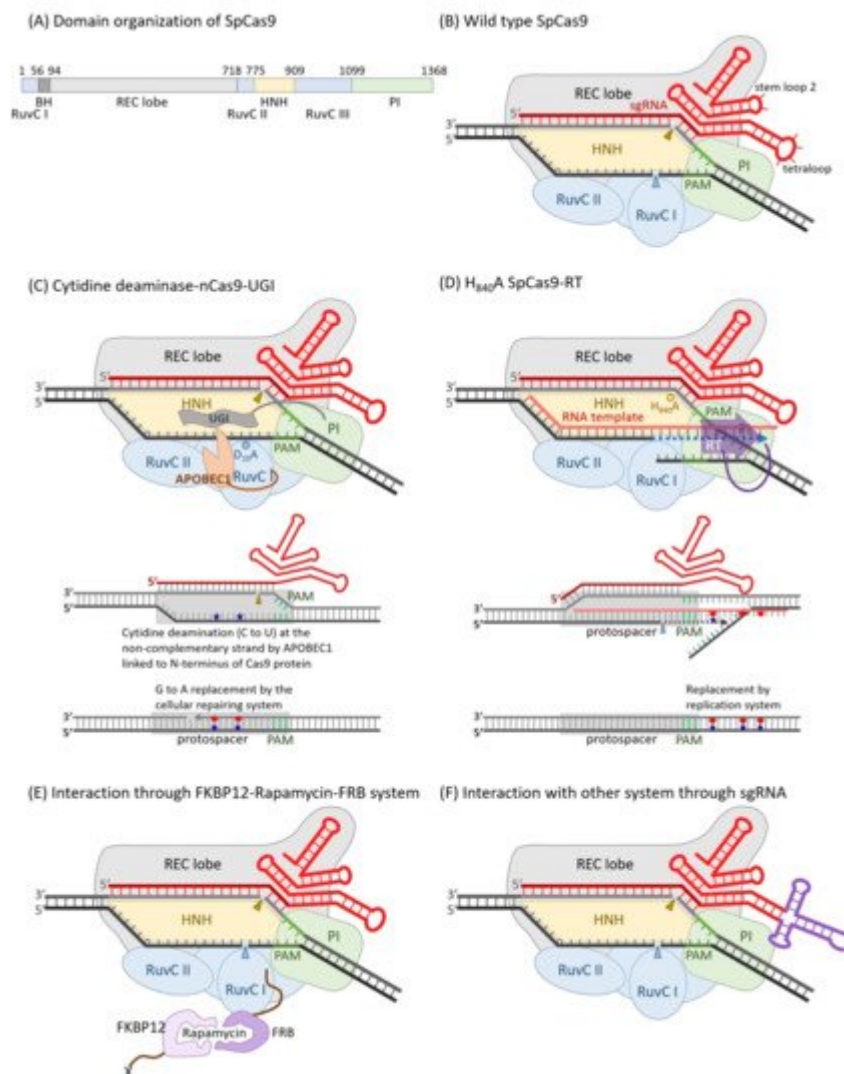


Figure 1. Diagrams of SpCas9 and its derivatives for various applications. The domain organization of SpCas9 (A) and a schematic diagram of wild-type SpCas9 associated with a sgRNA (B) was illustrated. The non-complementary strand is cut by the RuvC nuclease domain, and this nuclease activity was blocked in D₁₀A mutant. On the other hand, the complementary strand was digested by the HNH nuclease domain, and such nuclease activity was destroyed in H₈₄₀A mutant. (C) The D₁₀A mutant, also named Cas9 nickase (nCas9), was engineered as a C to T nucleotide editor by linking a cytidine deaminase, APOBEC1, on the N-terminus of it and the switching probability could be elevated by the fusion of a uracil glycosylase inhibitor (UGI) on the C-terminus of nCas9. Like TALE, dCas9 could be guided by a sgRNA as a sequence-specific DNA-binding riboprotein. Transcriptional regulators, DNA modification enzymes, or histone modification enzymes could be fused to either or both of the N- and C-termini. In case of reverse transcriptase, it was fused to the C-terminus of Cas9, accompanied by an RNA template with 3'-end complementary to the non-complementary strand of protospacer, which could alter the nearby

nucleotides downstream the RuvC cutting site (D). The localization of the Cas9-sgRNA also could be guided to an X-protein through an FKBP12–rapamycin–FRB bridge (E). The localization of the Cas9-sgRNA could also be guided via certain specific interactions, such as those between aptamer RNA and ABP. The tetraloop was replaced by an RNA aptamer of unique secondary structure, which can be recognized by a specific aptamer binding protein (F).

Table 1. The engineered mutations of SpCas9 on target site recognition and nuclease activities.

| Domain Where Mutations Engineered | | | | | | | Functional Changes | Reference |
|-----------------------------------|-------------------|-------------------|--|-------------------|--------------------|---------------------|--|---|
| Mutants | RuvC-I (1–55) | BH (56–93) | REC Lobe (94–717) | RuvC-II (718–774) | HNH (775–908) | RuvC-III (909–1098) | PI (1099–1368) | |
| SpG | | | | | | | D ₁₁₃₅ L, S ₁₁₃₆ W, G ₁₂₁₈ Q, E ₁₂₁₉ Q, R ₁₃₃₅ Q, T ₁₃₃₇ R | The recognition sequence of PAM changed from NGG to NGN. |
| SpRY | | A ₆₁ R | | | | | L ₁₁₁₁ R, D ₁₁₃₅ L, S ₁₁₃₆ W, G ₁₂₁₈ Q, E ₁₂₁₉ Q, N ₁₃₁₇ R, A ₁₃₂₂ R, R ₁₃₃₃ P, R ₁₃₃₅ Q, T ₁₃₃₇ R | The recognition sequence of PAM changed from NGG to PAMless (NRN > NYN). [69] |
| eSpCas9(1.1) | | | | | K ₈₄₈ A | | K ₁₀₀₃ A, R ₁₀₆₀ A | eSpCas9(1.1) displayed efficient and precise genome editing in human cells. [70] |
| SpCas9-HF1 | | | N ₄₉₇ A, R ₆₆₁ A, Q ₆₉₅ A | | | Q ₉₂₆ A | | SpCas9-HF1 performed with high on-target activity and reduced off-target editing. [71] [72] |
| nSpCas9 | D ₁₀ A | | | | | | | RuvC nuclease activity was eliminated. (nickase) [58] |

| Domain Where Mutations Engineered | | | | | | | Functional Changes | Reference |
|-----------------------------------|-------------------|---------------|----------------------|----------------------|--------------------|------------------------|--------------------|--|
| Mutants | RuvC-I (1–55) | BH (56–93) | REC Lobe (94–717) | RuvC-II (718–774) | HNH (775–908) | RuvC-III (909–1098) | PI (1099–1368) | |
| SpCas9(H ₈₄₀ A) | | | | | H ₈₄₀ A | | | HNH nuclease activity was diminished. |
| dSpCas9 | D ₁₀ A | | | | H ₈₄₀ A | | | Dead SpCas9 lost both of the RuvC and HNH nuclease activities. |
| SpCas9(N ₈₆₃ A) | | | | | N ₈₆₃ A | | | HNH nuclease activity was eliminated. [67] |
| SpCas9(D ₈₃₉ A) | | | | | D ₈₃₉ A | | | HNH nuclease activity was diminished. [66] |
| SpCas9(H ₉₈₃ A) | | | | | | H ₉₈₃ A | | RuvC nuclease activity was eliminated. [65] |

Besides Cas9, which recognized G-rich PAM at 3' end of protospacer, class 2 type V Cas12a (originally called Cpf1) effector enzymes also became attractive [73]. The long pre-crRNA was bound and processed by an intrinsic RNase activity of Cas12a protein to mature crRNA, which was composed of a repeat sequence at the 5' end and spacer at the 3' end. This characteristic was utilized to design multiple crRNA in a single RNA transcript [73][74]. A canonical TTTV PAM was at the 5' end of a 23 bp protospacer. Only a short 42–44 nt crRNA, which was composed of 19 nt repeat and 23–25 nt spacers, was necessary to guide the Cas12a's RNA-dependent endonuclease activity, of which DNA was cut at the PAM-distal end to leave 5' protruding staggered ends. Like Cas9, the RuvC nuclease domain was involved in non-complementary strand cleavage, while a new Nuc domain, instead of the HNH domain, was used in Cas12a for complementary strand cleavage [75]. The size of *Lachnospiraceae bacterium* MA2020 Cas12a (LbCas12a) was merely 1206 a.a. and as active as the most widely used Cas12a isolated from *Acidaminococcus* sp. (AsCas12a, 1307 a.a.). Engineered LbCas12a with Q₅₇₁K and C₁₀₀₃Y mutations, referred to as Lb2Cas12a, was more active and could recognize both TTTV and CTTV PAM motives [76].

References

1. Sansbury, B.M.; Kmiec, E.B. On the Origins of Homology Directed Repair in Mammalian Cells. *Int. J. Mol. Sci.* 2021, 22, 3348.

2. Ramirez, C.L.; Certo, M.T.; Mussolino, C.; Goodwin, M.J.; Cradick, T.J.; McCaffrey, A.P.; Cathomen, T.; Scharenberg, A.M.; Joung, J.K. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.* 2012, 40, 5560–5568.
3. Urnov, F.D.; Miller, J.C.; Lee, Y.-L.; Beausejour, C.M.; Rock, J.M.; Augustus, S.; Jamieson, A.C.; Porteus, M.H.; Gregory, P.D.; Holmes, M.C. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005, 435, 646–651.
4. Gregory, T.R.; Nicol, J.A.; Tamm, H.; Kullman, B.; Kullman, K.; Leitch, I.J.; Murray, B.G.; Kapraun, D.F.; Greilhuber, J.; Bennett, M.D. Eukaryotic genome size databases. *Nucleic Acids Res.* 2007, 35, D332–D338.
5. Redi, C.; Capanna, E. Genome size evolution: Sizing mammalian genomes. *Cytogenet. Genome Res.* 2012, 137, 97–112.
6. Bos, J.; Heyting, C.; Borst, P.; Arnberg, A.; Van Bruggen, E. An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA. *Nature* 1978, 275, 336–338.
7. Bolotin, M.; Coen, D.; Deutsch, J.; Dujon, B.; Netter, P.; Petrochilo, E.; Slonimski, P. Recombination of mitochondria in *saccharomyces-cerevisiae*. *Bull. de l'Inst. Pasteur* 1971, 69, 215–239.
8. Jacquier, A.; Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 1985, 41, 383–394.
9. Stoddard, B.L. Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure* 2011, 19, 7–15.
10. Stoddard, B.L. Homing endonuclease structure and function. *Q. Rev. Biophys.* 2005, 38, 49–95.
11. Taylor, G.K.; Stoddard, B.L. Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.* 2012, 40, 5189–5200.
12. Hafez, M.; Hausner, G. Homing endonucleases: DNA scissors on a mission. *Genome* 2012, 55, 553–569.
13. Pâques, F.; Duchateau, P. Meganucleases and DNA double-strand break-induced recombination: Perspectives for gene therapy. *Curr. Gene Ther.* 2007, 7, 49–66.
14. Li, H.; Pellenz, S.; Ulge, U.; Stoddard, B.L.; Monnat, R.J., Jr. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res.* 2009, 37, 1650–1662.
15. Arnould, S.; Delenda, C.; Grizot, S.; Desseaux, C.; Paques, F.; Silva, G.; Smith, J. The I-Crel meganuclease and its engineered derivatives: Applications from cell modification to gene therapy. *Protein Eng. Des. Sel.* 2011, 24, 27–31.

16. Li, L.; Wu, L.P.; Chandrasegaran, S. Functional domains in Fok I restriction endonuclease. *Proc. Natl. Acad. Sci. USA* 1992, 89, 4275–4279.
17. Kim, Y.-G.; Chandrasegaran, S. Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci. USA* 1994, 91, 883–887.
18. Kim, Y.-G.; Cha, J.; Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* 1996, 93, 1156–1160.
19. Persikov, A.V.; Wetzel, J.L.; Rowland, E.F.; Oakes, B.L.; Xu, D.J.; Singh, M.; Noyes, M.B. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res.* 2015, 43, 1965–1984.
20. Chandrasegaran, S.; Smith, J. Chimeric restriction enzymes: What is next? *Biol. Chem.* 1999, 380, 841–848.
21. Smith, J.; Bibikova, M.; Whitby, F.G.; Reddy, A.; Chandrasegaran, S.; Carroll, D. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.* 2000, 28, 3361–3369.
22. Bibikova, M.; Carroll, D.; Segal, D.J.; Trautman, J.K.; Smith, J.; Kim, Y.-G.; Chandrasegaran, S. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* 2001, 21, 289–297.
23. Şöllü, C.; Pars, K.; Cornu, T.I.; Thibodeau-Beganny, S.; Maeder, M.L.; Joung, J.K.; Heilbronn, R.; Cathomen, T. Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res.* 2010, 38, 8269–8276.
24. Doyon, Y.; Vo, T.D.; Mendel, M.C.; Greenberg, S.G.; Wang, J.; Xia, D.F.; Miller, J.C.; Urnov, F.D.; Gregory, P.D.; Holmes, M.C. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* 2011, 8, 74–79.
25. Persikov, A.V.; Singh, M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* 2014, 42, 97–108.
26. Garton, M.; Najafabadi, H.S.; Schmitges, F.W.; Radovani, E.; Hughes, T.R.; Kim, P.M. A structural approach reveals how neighbouring C2H2 zinc fingers influence DNA binding specificity. *Nucleic Acids Res.* 2015, 43, 9147–9157.
27. Bonas, U.; Stall, R.E.; Staskawicz, B. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet. MGG* 1989, 218, 127–136.
28. Deng, W.; Marshall, N.C.; Rowland, J.L.; McCoy, J.M.; Worrall, L.J.; Santos, A.S.; Strynadka, N.C.; Finlay, B.B. Assembly, structure, function and regulation of type III secretion systems. *Nat. Rev. Microbiol.* 2017, 15, 323–337.

29. Dey, S.; Chakravarty, A.; Guha Biswas, P.; De Guzman, R.N. The type III secretion system needle, tip, and translocon. *Protein Sci.* 2019, 28, 1582–1593.
30. Boch, J.; Bonas, U. Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.* 2010, 48, 419–436.
31. Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009, 326, 1509–1512.
32. Deng, D.; Yan, C.; Pan, X.; Mahfouz, M.; Wang, J.; Zhu, J.-K.; Shi, Y.; Yan, N. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 2012, 335, 720–723.
33. Bradley, P. Structural modeling of TAL effector–DNA interactions. *Protein Sci.* 2012, 21, 471–474.
34. Moore, R.; Chandrabhas, A.; Bleris, L. Transcription activator-like effectors: A toolkit for synthetic biology. *ACS Synth. Biol.* 2014, 3, 708–716.
35. Streubel, J.; Blücher, C.; Landgraf, A.; Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 2012, 30, 593–595.
36. Richter, A.; Streubel, J.; Boch, J. TAL effector DNA-binding principles and specificity. In *TALENs*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 9–25.
37. Carlson, D.F.; Tan, W.; Lillico, S.G.; Stverakova, D.; Proudfoot, C.; Christian, M.; Voytas, D.F.; Long, C.R.; Whitelaw, C.B.A.; Fahrenkrug, S.C. Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. USA* 2012, 109, 17382–17387.
38. Miller, J.C.; Tan, S.; Qiao, G.; Barlow, K.A.; Wang, J.; Xia, D.F.; Meng, X.; Paschon, D.E.; Leung, E.; Hinkley, S.J. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 2011, 29, 143–148.
39. Mussolino, C.; Morbitzer, R.; Lütge, F.; Dannemann, N.; Lahaye, T.; Cathomen, T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 2011, 39, 9283–9293.
40. Schreiber, T.; Sorgatz, A.; List, F.; Blüher, D.; Thieme, S.; Wilmanns, M.; Bonas, U. Refined requirements for protein regions important for activity of the TALE AvrBs3. *PLoS ONE* 2015, 10, e0120214.
41. Gao, H.; Wu, X.; Chai, J.; Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* 2012, 22, 1716–1720.
42. Doyle, E.L.; Stoddard, B.L.; Voytas, D.F.; Bogdanove, A.J. TAL effectors: Highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol.* 2013, 23, 390–398.

43. Cong, L.; Zhou, R.; Kuo, Y.-c.; Cunniff, M.; Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.* 2012, 3, 1–6.
44. Cermak, T.; Doyle, E.L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J.A.; Somia, N.V.; Bogdanove, A.J.; Voytas, D.F. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 2011, 39, e82.
45. Cermak, T.; Starker, C.G.; Voytas, D.F. Efficient design and assembly of custom TALENs using the Golden Gate platform. In *Chromosomal Mutagenesis*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 133–159.
46. Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* 2011, 39, 5790–5799.
47. Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. Assembly of designer TAL effectors by Golden Gate cloning. *PLoS ONE* 2011, 6, e19722.
48. Zhang, F.; Cong, L.; Lodato, S.; Kosuri, S.; Church, G.M.; Arlotta, P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* 2011, 29, 149–153.
49. Deng, P.; Carter, S.; Fink, K. Design, Construction, and Application of Transcription Activation-Like Effectors. In *Viral Vectors for Gene Therapy*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 47–58.
50. Nitsch, S.; Mussolino, C. Generation of TALE-based designer epigenome modifiers. In *Epigenome Editing*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 89–109.
51. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2005, 151, 2551–2561.
52. Pourcel, C.; Salvignol, G.; Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 2005, 151, 653–663.
53. Hille, F.; Charpentier, E. CRISPR-Cas: Biology, mechanisms and relevance. *Philos. Trans. R. Soc. B Biol. Sci.* 2016, 371, 20150496.
54. Hille, F.; Richter, H.; Wong, S.P.; Bratovič, M.; Ressel, S.; Charpentier, E. The biology of CRISPR-Cas: Backward and forward. *Cell* 2018, 172, 1239–1259.
55. Koonin, E.V.; Makarova, K.S.; Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 2017, 37, 67–78.
56. Koonin, E.V.; Makarova, K.S. Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. B* 2019, 374, 20180087.

57. Makarova, K.S.; Wolf, Y.I.; Iranzo, J.; Shmakov, S.A.; Alkhnbashi, O.S.; Brouns, S.J.; Charpentier, E.; Cheng, D.; Haft, D.H.; Horvath, P. Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 2020, 18, 67–83.
58. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012, 337, 816–821.
59. Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014, 513, 569–573.
60. Jinek, M.; Jiang, F.; Taylor, D.W.; Sternberg, S.H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 2014, 343.
61. Josephs, E.A.; Kocak, D.D.; Fitzgibbon, C.J.; McMenemy, J.; Gersbach, C.A.; Marszalek, P.E. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* 2015, 43, 8924–8941.
62. Sternberg, S.H.; LaFrance, B.; Kaplan, M.; Doudna, J.A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* 2015, 527, 110–113.
63. Jiang, F.; Doudna, J.A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* 2017, 46, 505–529.
64. Le Rhun, A.; Escalera-Maurer, A.; Bratovič, M.; Charpentier, E. CRISPR-Cas in *Streptococcus pyogenes*. *RNA Biol.* 2019, 16, 380–389.
65. Furuhashi, Y.; Kato, Y. Asymmetric Roles of Two Histidine Residues in *Streptococcus pyogenes* Cas9 Catalytic Domains upon Chemical Rescue. *Biochemistry* 2021, 60, 194–200.
66. Zuo, Z.; Zolekar, A.; Babu, K.; Lin, V.J.; Hayatshahi, H.S.; Rajan, R.; Wang, Y.-C.; Liu, J. Structural and functional insights into the bona fide catalytic state of *Streptococcus pyogenes* Cas9 HNH nuclease domain. *Elife* 2019, 8, e46500.
67. Nishimasu, H.; Ran, F.A.; Hsu, P.D.; Konermann, S.; Shehata, S.I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014, 156, 935–949.
68. Kleinstiver, B.P.; Prew, M.S.; Tsai, S.Q.; Topkar, V.V.; Nguyen, N.T.; Zheng, Z.; Gonzales, A.P.; Li, Z.; Peterson, R.T.; Yeh, J.-R.J. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 2015, 523, 481–485.
69. Walton, R.T.; Christie, K.A.; Whittaker, M.N.; Kleinstiver, B.P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 2020, 368, 290–296.

70. Slaymaker, I.M.; Gao, L.; Zetsche, B.; Scott, D.A.; Yan, W.X.; Zhang, F. Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016, 351, 84–88.
71. Chen, J.S.; Dagdas, Y.S.; Kleinstiver, B.P.; Welch, M.M.; Sousa, A.A.; Harrington, L.B.; Sternberg, S.H.; Joung, J.K.; Yildiz, A.; Doudna, J.A. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* 2017, 550, 407–410.
72. Kleinstiver, B.P.; Pattanayak, V.; Prew, M.S.; Tsai, S.Q.; Nguyen, N.T.; Zheng, Z.; Joung, J.K. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016, 529, 490–495.
73. Zetsche, B.; Gootenberg, J.S.; Abudayyeh, O.O.; Slaymaker, I.M.; Makarova, K.S.; Essletzbichler, P.; Volz, S.E.; Joung, J.; Van Der Oost, J.; Regev, A. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015, 163, 759–771.
74. Zhong, G.; Wang, H.; Li, Y.; Tran, M.H.; Farzan, M. Cpf1 proteins excise CRISPR RNAs from mRNA transcripts in mammalian cells. *Nat. Chem. Biol.* 2017, 13, 839–841.
75. Yamano, T.; Nishimasu, H.; Zetsche, B.; Hirano, H.; Slaymaker, I.M.; Li, Y.; Fedorova, I.; Nakane, T.; Makarova, K.S.; Koonin, E.V. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 2016, 165, 949–962.
76. Tran, M.H.; Park, H.; Nobles, C.L.; Karunadharma, P.; Pan, L.; Zhong, G.; Wang, H.; He, W.; Ou, T.; Crynen, G. A more efficient CRISPR-Cas12a variant derived from Lachnospiraceae bacterium MA2020. *Mol. Ther. -Nucleic Acids* 2021, 24, 40–53.

Retrieved from <https://www.encyclopedia.pub/entry/history/show/34573>