Cell PRESS

RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing?

Giedrius Gasiunas and Virginijus Siksnys

Institute of Biotechnology, Vilnius University, Graiciuno 8, LT-02241, Vilnius, Lithuania

Tailor-made nucleases for precise genome modification, such as zinc finger or TALE nucleases, currently represent the state-of-the-art for genome editing. These nucleases combine a programmable protein module which guides the enzyme to the target site with a nuclease domain which cuts DNA at the addressed site. Reprogramming of these nucleases to cut genomes at specific locations requires major protein engineering efforts. RNA-guided DNA endonuclease Cas9 of the type II (clustered regularly interspaced short palindromic repeat) CRISPR–Cas system uses CRISPR RNA (crRNA) as a guide to locate the DNA target and the Cas9 protein to cut DNA. Easy programmability of the Cas9 endonuclease using customizable RNAs brings unprecedented flexibility and versatility for targeted genome modification. We highlight the potential of the Cas9 RNA-guided DNA endonuclease as a novel tool for genome surgery, and discuss possible constraints and future prospects.

Gene editing

Targeted genome editing technology that enables the generation of site-specific changes in the genomic DNA of cellular organisms is a Holy Grail for genome engineers [1–3]. Currently available genome editing technologies rely on the double-strand break (DSB) repair pathways of the cell. When a DSB occurs in DNA, it triggers a natural process of DNA repair either by an 'error-prone' non-homologous end joining (NHEJ; see Glossary) [4] or by homologous recombination (HR) [5]. Therefore, molecular tools that can generate DSBs at specific sites within the genome are at the core of current genome editing technologies. The ideal gene editing tool should meet the following criteria: (i) high frequency of desired sequence changes in the target cell population; (ii) no off-target cleavage; and (iii) rapid and efficient assembly of nucleases that target any site on the genome at low cost [1].

Engineered, highly specific DNA endonucleases (meganucleases) programmable according to the desired specificity are currently the state-of-the-art tools for gene editing technology. Meganucleases are by definition sequence-specific endonucleases with large (>12 bp) recognition sites [6]. Meganucleases combine a programmable specificity module which guides the enzyme to the target site with a nuclease domain which introduces cleavage at the addressed site. Different protein scaffolds are currently used for the development of meganucleases for gene targeting (Table 1): (i) homing endonucleases (HEases) [6–8]; (ii) zinc finger nucleases (ZFNs) [9]; (iii) TALE nucleases (TALENs) [10,11]; and (iv) restriction enzyme-triple helix forming oligonucleotide conjugates (RE-TFOs) [12,13].

Naturally occurring HEases or their engineered variants, which recognize long DNA sequences (up to 40 base

Glossary

cas: CRISPR-associated genes which are located in the vicinity of CRISPR array and are necessary for the silencing of invading nucleic acid.

Cas9t: Cas9–crRNA–tracrRNA ternary complex, which functions as an RNA-guided DNA endonuclease and mediates site-specific DNA cleavage.

Clustered regularly interspaced short palindromic repeat (CRISPR): an array of short conserved repeat sequences interspaced by unique DNA sequences of similar size called spacers, which often originate from phage or plasmid DNA. CRISPR array together with *cas* genes form the CRISPR–Cas system, which functions as an adaptive immune system in prokaryotes.

CRISPR RNA (crRNA): small RNA molecule generated by transcription and processing of the CRISPR array. crRNA is composed of a conserved repeat fragment(s) and a variable spacer sequence, which matches the complimentary sequence in the invading nucleic acid.

Homologous repair (HR): error-free DNA repair pathway that seals the broken DNA molecule using a homologous sequence (template).

Non-homologous end joining (NHEJ): a pathway that repairs DNA DSBs in the absence of a homologous template; usually leads to small insertions or deletions.

Protospacer adjacent motif (PAM): a short conserved nucleotide stretch located in the vicinity of a protospacer in the target DNA and necessary for DNA cleavage by Cas9t.

Protospacer: a fragment in the target DNA, which matches a spacer sequence in the CRISPR array.

Single guide RNA (sgRNA): RNA hairpin obtained by connecting crRNA and tracrRNA into a single molecule.

Transcription activator-like effector nuclease (TALEN): an artificial nuclease obtained by fusing *Xanthomonas* transcription activator-like effector (TALE) DNA binding domains to the nonspecific nuclease domain.

Trans-acting CRISPR RNA (tracrRNA): trans-encoded small RNA molecule, which forms a duplex with a repeat fragment of crRNA.

Triple helix forming oligonucleotide (TFO): an artificial oligodeoxynucleotide, which binds to the polypurine sequences of the double-stranded DNA forming DNA triple helix.

Zinc finger nuclease (ZFN): an artificial nuclease created by fusing zinc finger motifs, which serve as DNA recognition modules, to a nonspecific DNA cleavage domain of the *Fok*I restriction endonuclease.

Corresponding author: Siksnys, V. (siksnys@ibt.lt).

Keywords: molecular tools; gene targeting; RNA-guided DNA cleavage; clustered regularly interspaced short palindromic repeat (CRISPR).

0966-842X/$ – see front matter

 http://dx.doi.org/10.1016/j.tim.2013.09.001

Table 1. Tools for gene editing

Tool	Specificity module	Cleavage module	Target site length, bp	Reprogramming	Targeting frequency	Specificity	Other features
HEase	Target recognition domain	Nuclease domain	14–40	Complicated: requires protein engineering	Low	Off-targeting reported	High cleavage efficiency
ZFN	Zinc finger domains	*Fok*I nuclease domain	18–36	Complicated: requires domain shuffling, assembly, and protein engineering, from 10 weeks[a]	High	Off-targeting reported	Sequence bias, some variants show toxicity
TALEN	TALE domains	*Fok*I nuclease domain	24–59	Relatively easy: requires domain shuffling, assembly, and protein engineering, from 4 weeks[b]	High, nearly every sequence	Systematically not determined	Large protein size
TFO conjugate	TFO (+restriction endonuclease)	Restriction endonuclease	4–8 + TFO	Relatively easy: requires only DNA oligo but includes a complicated chemistry step	Low, restricted by RE and TFO sequences	Systematically not determined	Slow equilibrium
Cas9t	RNA (+ PAM)	Cas9	20 + PAM (2–5)	Easy and fast: requires only sgRNA	High, depends on PAM	Off-targeting reported	Multiplexing possibilities

[a]According to manufacturer's information (http://www.sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/custom-zfn.html).

[b]According to manufacturer's information (http://www.cellectis-bioresearch.com/products/talen-basic).

pairs), have been applied to gene editing in numerous experimental designs and cell types [6–8]. Artificial nucleases, such as ZFNs or TALENs, have been developed for the purpose of gene targeting. ZFNs are created by fusing zinc finger motifs, which serve as a DNA recognition module, to a nonspecific DNA cleavage domain of the *Fok*I restriction endonuclease [9]. The combination of zinc finger modules specific for different sequences enables targeting of desired sites within the genome. TALENs are based on the fusion of the *Xanthomonas* transcription activator-like effector (TALE) DNA-binding domain (DBD) to the nuclease [10,11]. The TALE DBD contains repeated motifs that recognize specific nucleotides, and therefore the specificity of TALENs could be programmed by selecting and shuffling repeat segments specific for different nucleotides. Despite the differences in protein scaffolds, specificity of HEases, ZFNs, and TALENs is determined by the DBD of the meganuclease. In an alternative approach, site-specific genome cleavage reagents have been engineered using DNA triple helix forming oligonucleotides (TFOs) as address tags fused to various DNA damaging compounds [14,15] or restriction enzyme nuclease modules [12,13].

HEases, ZFNs, and TALENs are already commercially exploited as molecular tools for the generation of DSBs at specific sites within chromosomes. One major limitation is that HEases, ZFNs, and TALENs must be re-engineered for each new DNA target. Despite progress in the field, engineering of the DBDs of ZFNs and TALENs is time consuming, requires considerable skills, and often produces variants that are poised for off-target cleavage [1,16]. RE-TFO gene targeting technology employs DNA triple helix forming nucleotides to achieve binding specificity [12–15]. In principle, TFO as a specificity module has an advantage because reprogramming requires only oligonucleotide synthesis. However, several drawbacks limit the application of this technology. First, oligonucleotide conjugation to the nuclease module requires complicated chemistry steps [12,13]. Second, triple helix formation is slow and often requires modified oligonucleotides to achieve desired target specificity [17]. Therefore, new genome editing tools, which are more affordable and easier and faster to engineer, are still in demand.

RNA-dependent DNA endonuclease Cas9 of the CRISPR system

CRISPR (clustered regularly interspaced short palindromic repeat) is a recently discovered bacterial and archaeal adaptive immune system, which consists of an array of short conserved repeat sequences interspaced by unique DNA sequences of similar size called spacers, which often originate from phage or plasmid DNA. CRISPR arrays together with *cas* (CRISPR-associated) genes form the CRISPR–Cas adaptive immune system [18–20]. CRISPR–Cas systems function by incorporating fragments of the invading nucleic acid as spacers into a host genome and later use these spacers as templates to generate small RNA molecules (crRNA) that are combined with Cas proteins into an effector complex which silences foreign nucleic acids in the subsequent rounds of infection. CRISPR–Cas systems are distinct and have been categorized into three main types, based on core element content and sequences [21]. The effector complex that binds crRNA and triggers cleavage of invading nucleic acid differs strikingly between different CRISPR subtypes. In type I systems, crRNAs are incorporated into a multi-subunit effector complex called Cascade (CRISPR-associated complex for antiviral defense), which binds to the target DNA and triggers degradation by an accessory Cas3 protein [22,23]. In type III CRISPR–Cas systems, exemplified by *Sulfolobus solfataricus* and *Pyrococcus furiosus*, the Cas RAMP module (Cmr) and crRNA complex recognize and cleave synthetic RNA *in vitro* [24,25]. Surprisingly, in type II CRISPR–Cas systems only the Cas9 protein (previously named Cas5 or Csn1) is required for DNA interference [19,26–28]. Cas9 is a large, multi-domain protein which contains two nuclease domains, an RuvC-like nuclease domain near the amino terminus and a HNH-like nuclease domain in the middle of the protein [29]. Cas9 forms a ternary complex (Cas9t) with two RNA molecules: crRNA