Computational and Structural Biotechnology Journal xxx (2017) xxx–xxx

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering

Tuhin Kumar Guha, Alvan Wai, Georg Hausner *

Department of Microbiology, University of Manitoba, Winnipeg, Manitoba R3T2N2, Canada

ARTICLE INFO

Article history:
Received 12 October 2016
Received in revised form 24 December 2016
Accepted 27 December 2016
Available online xxxx

Keywords:
Meganuclease
TALEN
Zinc finger nuclease
CRISPR/Cas9
Regulatory switch
Hammerhead ribozyme

ABSTRACT

Targeted genome editing has become a powerful genetic tool for studying gene function or for modifying genomes by correcting defective genes or introducing genes. A variety of reagents have been developed in recent years that can generate targeted double-stranded DNA cuts which can be repaired by the error-prone, non-homologous end joining repair system or via the homologous recombination-based double-strand break repair pathway provided a suitable template is available. These genome editing reagents require components for recognizing a specific DNA target site and for DNA-cleavage that generates the double-stranded break. In order to reduce potential toxic effects of genome editing reagents, it might be desirable to control the in vitro or in vivo activity of these reagents by incorporating regulatory switches that can reduce off-target activities and/or allow for these reagents to be turned on or off. This review will outline the various genome editing tools that are currently available and describe the strategies that have so far been employed for regulating these editing reagents. In addition, this review will examine potential regulatory switches/strategies that can be employed in the future in order to provide temporal control for these reagents.

Contents

1. Introduction

One of the challenges in biotechnology has been developing efficient and reliable ways to make targeted changes within the genome of cells. Traditional approaches of mutagenesis utilizing chemical agents or transposons can require extensive screening in order to recover desired mutations [1–6]. Genome editing strategies using double-stranded (ds) DNA viral vectors in differentiated human cells and RNA interference (RNAi) mediated targeted gene knockdown approaches also have some pitfalls [7–10]. For example, the protein composition of the viral capsid can be potentially immunogenic. Moreover, abnormal gene expression along with insertional mutagenesis may be triggered if there are random mutations in the viral sequences. On the other hand, the use of exogenously introduced dsRNA in RNAi technology can disrupt the "homeostasis" of the cellular machinery involved in gene silencing. Currently, the most popular genome engineering techniques apply DNA-cutting enzymes/complexes that generate targeted double-strand cuts [11–13], which are repaired by the host cells by either the error-prone, non-homologous end joining repair system (NHEJ), or the homologous recombination-based double-strand break repair pathway (HDR) [14–18]. The most frequent application of these

* Corresponding author.
E-mail address: hausnerg@cc.umanitoba.ca (G. Hausner).

http://dx.doi.org/10.1016/j.csbj.2016.12.006

Please cite this article as: Guha TK, et al, Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering, Comput Struct Biotechnol J (2017), http://dx.doi.org/10.1016/j.csbj.2016.12.006

endonuclease-based tools is the study of gene function through the inactivation of the target gene [19–21]. In addition, by providing a repair template, these systems allow for gene replacement strategies by taking advantage of the host cell's dsDNA break homologous repair system [22–24]. These new methods have tremendous potential towards the development of more accurate cellular and humanized laboratory animal models for various pathological conditions [25,26]. Moreover, these endonuclease-based genetic engineering techniques are being developed as therapeutic agents to cure human monogenic diseases [27–31]. Genome editing tools have far-reaching implications in the agricultural sector and in their potential of curbing pest populations, such as malaria insect vectors, or invasive species, such as cane toads and carps [32–36]. The latter applications are achieved in promoting the 'gene drive' of an introduced genetic element (such as a meganuclease) within an interbreeding population that can distort sex ratios (daughterless generations), or target genes related to fertility or pathogenicity [37–41].

Genome editing tools include meganucleases (MNs) [42–45], zinc finger nucleases (ZFNs) [46–49], transcription activator-like effector nucleases (TALENs) [50–53], clustered regularly interspaced short palindromic repeat (CRISPR)-associated nuclease Cas9 [54–56], and targetrons [57–63]. All of them can achieve precise genetic modifications by inducing targeted DNA double-strand breaks (DSBs). Depending on the cell cycle stage, as well as the presence or absence of a repair template with homologous terminal regions, the DSB may then be repaired by either NHEJ or HDR [64–68]. NHEJ can result in frameshift mutations that usually lead to gene disruption or gene knockout and/or the production of nonfunctional truncated proteins [69–71]; one exception being when a frameshift mutation was introduced to correct a defective coding sequence in the dystrophin gene [72,73]. In contrast, when single- or double-stranded DNA templates with homologous sequences that correspond to sequences flanking the break site are introduced within the cell, the lesion may be repaired using the HDR machinery [74,75].

One crucial concern when applying these genetic editing tools is the potential of cleavage at non-targeted sites. This event can be lethal or generate undesirable mutations resulting in the requirement of extensive screening in order to identify cells with the desired site-specific modifications. Many excellent reviews are available with regards to the above listed genome editing tools [13,21,42,44,45,76–87]. Therefore, this review will provide only a brief overview of the current genome editing tools and note any modifications made within recent years. The major focus in this review is to examine the efforts that have been made in the development of programmable, endonuclease-based platforms and various molecular switches that could be employed for the temporal regulation of these DNA-cutting enzymes in order to reduce off-target activities. The term "programmable" refers to the ability to engineer the nuclease-based platforms for recognizing various target sites (i.e. target specificity) in the genome.

2. Genome Editing Reagents

In general, genome editing tools using DSB nuclease-driven reactions (Fig. 1) can be divided into two groups. The first group consists of MNs, ZFNs and TALENs, which achieve sequence-specific DNA-binding via protein-DNA interactions [13,42]. The second group is comprised of two sub-groups: (i) CRISPR/Cas9 and targetrons, which are RNA-guided systems [56,57] and (ii) peptide nucleic acids (PNAs), triplex-forming oligonucleotides (TFOs), and structure-guided endonucleases (SGNs), which are DNA-based-guided systems [88–92]. A generalized comparison for the more commonly used genome engineering tools is presented in Table 1.

Meganucleases, or homing endonucleases (HEases; Fig. 1a,b), are highly site-specific dsDNA endonucleases that can be reengineered to expand their target site repertoires using various strategies, such as computational structure-based design, domain swapping, combined with yeast surface display for efficient detection of HEases with desired sequence specificities [93–98]. The LAGLIDADG family of MNs have been extensively studied and applied as genome editing tools [43,44,45,99–101]. Unless otherwise mentioned, we are referring to LAGLIDADG enzymes as MNs for simplicity. One essential drawback for this class of enzyme is its non-modular configuration. The DNA recognition and cleavage functions can be, in part, intertwined in a single protein domain. Therefore, engineering of MNs has been challenging [45,76] and has resulted in the development of other editing tools. However, a recent study suggests that there are multiple points across the LAGLIDADG protein that can be involved in holding metal ions in suitable positions to facility cleavage [102]. This finding along with technologies, such as yeast surface display-SELEX, still hold promise for MNs to be engineered more efficiently in the near future [97]. Moreover, a single-chain modular nuclease architecture, termed 'megaTAL' (Fig. 1c), was designed in which the DNA-binding region of a transcription activator-like (TAL) effector is appended to a site-specific MN for cleaving a desired genomic target site [103]. The latter synthetic version of a MN provides a modular design, separating the endonuclease and DNA binding activities. Therapeutic applications that demand precision with regards to gene modification activity can be addressed by these engineered variants of MNs, as they are considered to be highly target-specific 'molecular scissors' [45]. MNs are also in demand as components of vector/cloning systems (e.g. HomeRun vector assembly system) and synthetic biology applications (e.g. iBrick) that require rare-cutting enzymes [104,105].

Even though the NHEJ pathway is usually exploited to introduce mutations at the DSBs within the genome [15,106], sometimes, DSBs possess compatible "sticky" ends that can be repaired without any introduced mutation [107]. Recently, the 'MegaTev' (Fig. 1d) architecture has been generated which involves fusion of the DNA-binding and cutting domain from a meganuclease (Mega, I-OnuI) with another nuclease domain derived from the GIY-YIG HEase (Tev, I-TevI). This protein was designed to position the two cutting domains ~30 bp apart on the DNA substrate and generate two DSBs with non-compatible single-stranded overhangs for more efficient gene disruption [108]. More recently, similar to the MegaTev concept, Wolfs et al.have designed another dual nuclease, in which the Tev endonuclease domain is attached to the Cas9 nuclease domain, known as TevCas9 [109]. This hybrid nuclease, when introduced within human embryonic kidney cells (HEK293) along with appropriate guide RNAs, has been shown to delete 33 to 36 bp of the target site, thereby creating two non-compatible DNA breaks at moderately higher frequencies (40%). Therefore, this newly designed dual active endonuclease also promises to favor genome editing events (i.e. introduce mutations) by avoiding the creation of compatible "sticky" ends which lead to a failed attempt of genome editing [109].

More recently developed genome editing tools try to be more flexible with regards to retargeting the reagent to different sequences by having a modular design: a DNA-cutting domain (that can be non-specific) and a distinct programmable DNA-binding domain. The ZFNs are artificial endonucleases that have been generated by combining a small zinc finger (ZF; ~30 amino acids) DNA-binding/recognition domain (Cys_2His_2) to a type IIS nonspecific DNA-cleavage domain from the FokI restriction enzyme (Fig. 1e). However, the cleavage activity of the FokI endonuclease demands dimerization [46,110]. As a ZF module recognizes a 3 bp sequence, there is a requirement for multiple fingers in each ZFN monomer for recognizing and binding to longer DNA target sequences [46]. In the past, using structure-based design, two ZFN variants were engineered that efficiently cleaved DNA only when paired as a heterodimer, thereby providing a potential avenue for improving the specificity of ZFNs as gene modification reagents [111]. In a different structure-based study, using 3D protein modeling and energy calculations through computer-based softwares, researchers have identified potential residues within the FokI dimer interface that are responsible for ZFN dimerization [112]. These newly designed ZFNs were considered significantly less genotoxic (i.e. cleavage at on-target sites) in the cell-based recombination studies because the homodimerization could be prevented

Please cite this article as: Guha TK, et al, Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering, Comput Struct Biotechnol J (2017), http://dx.doi.org/10.1016/j.csbj.2016.12.006