Available online at www.sciencedirect.com
SciVerse ScienceDirect
Journal of Genetics and Genomics 40 (2013) 271−279
JGG

REVIEW

TALENs: Customizable Molecular DNA Scissors for Genome Engineering of Plants

Kunling Chen, Caixia Gao*

State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Received 18 February 2013; revised 18 March 2013; accepted 18 March 2013
Available online 26 March 2013

ABSTRACT

Precise genome modification with engineered nucleases is a powerful tool for studying basic biology and applied biotechnology. Transcription activator-like effector nucleases (TALENs), consisting of an engineered specific (TALE) DNA binding domain and a *Fok* I cleavage domain, are newly developed versatile reagents for genome engineering in different organisms. Because of the simplicity of the DNA recognition code and their modular assembly, TALENs can act as customizable molecular DNA scissors inducing double-strand breaks (DSBs) at given genomic location. Thus, they provide a valuable approach to targeted genome modifications such as mutations, insertions, replacements or chromosome rearrangements. In this article, we review the development of TALENs, and summarize the principles and tools for TALEN-mediated gene targeting in plant cells, as well as current and potential strategies for use in plant research and crop improvement.

KEYWORDS: TALENs; Genome engineering; Targeted gene modification; Plant

INTRODUCTION

Targeted genome modification (TGM), mediated by engineered nucleases, has been widely used to investigate gene function and expand biotechnology applications in yeast (Scherer and Davis, 1979), fruit fly (Bibikova et al., 2002), mice (Capecchi, 2005), human cell lines (Urnov et al., 2005), plants (Wright et al., 2005; de Pater et al., 2009) and many other organisms (Remy et al., 2010; Urnov et al., 2010). Reliable and efficient methods for obtaining site-specific modifications in plants are long sought-after goals for basic plant research and crop improvement (Pennisi, 2010). Although first demonstrated in the late 1980's, TGM in plants was far from routine because of its extremely low efficiency (typically about 10^{-5}) (Paszkowski et al., 1988; Halfter et al., 1992). The critical step in TGM is the introduction of DNA double-strand breaks (DSBs) at given genomic sites. It was soon discovered that engineered nucleases could generate DSBs and consequently activate DNA repair to seal the breaks along with any modifications such as mutations, insertions, replacements, and chromosomal rearrangements (Rouet et al., 1994).

Traditionally, TGM was mostly performed using zinc finger nucleases (ZFNs), artificial nucleases that consist of a synthetic ZFN domain fused to a *Fok* I cleavage domain (Urnov et al., 2010; Carroll, 2011). ZFNs have been used to modify endogenous genes in a wide range of organisms and cell types, and plant species including *Arabidopsis* (Osakabe et al., 2010; Zhang et al., 2010), tobacco (Townsend et al., 2009), maize (Shukla et al., 2009) and soybean (Curtin et al., 2011). Major constraints on ZFN application include the limited number of available target sites, more context dependence effects between the repeat units, low targeting efficiency and specificity, and

* Corresponding author. Tel/fax: +86 10 6480 7727.
E-mail address: cxgao@genetics.ac.cn (C. Gao).

http://dx.doi.org/10.1016/j.jgg.2013.03.009

frequent off-target effects caused partly by nonspecific DNA binding (DeFrancesco, 2011).

More recently, transcription activator-like effector nucleases (TALENs) have emerged as alternatives to ZFNs for TGM, and have been shown to have great potential for precise genome manipulation (Christian et al., 2010). Like ZFNs, TALENs consist of an engineered specific TALE DNA binding domain and a *Fok* I cleavage domain. The customizable TALE (Transcription activator-like effectors) DNA binding domain, composed of several nearly identical tandem repeat arrays, can target any given sequence according to a simple repeat variable di-residue (RVD)-nucleotide recognition code (Bogdanove et al., 2010; Bogdanove and Voytas, 2011). Within the past three years TALEN-mediated genome modification has been widely adopted in yeast (Li et al., 2011), nematode (Wood, 2011), fruit fly (Liu et al., 2012), rat (Tesson, 2011; Tong et al., 2012), human somatic and pluripotent cells (Hockemeyer et al., 2011; Miller et al., 2011), silkworm (Ma et al., 2012), livestock (Carlson et al., 2012), plants (Li et al., 2012b; Mahfouz et al., 2011; Shan et al., 2013; Zhang et al., 2013), *Xenopus* embryo (Lei, 2012), zebrafish (Huang et al., 2011; Sander et al., 2011; Bedell, 2012; Cade et al., 2012; Dahlem et al., 2012; Moore, 2012) and many other organisms, and is recognized as a major breakthrough in the core technology of genome engineering. Genome editing with engineered nucleases was named 2011 "Method of Year" (2011), and one year later genome engineering using TALENs and Cas9 was crowned 2012 "Breakthrough of the Year" by Science (2012).

Here we review the DNA targeting characteristics of TALENs and describe the available TALEN construction platforms. We also summarize the applications that have been made in plants, and discuss future work to enhance the utility of these tools in plant research and crop improvement.

DNA BINDING SPECIFICITY

TALEs are major virulence factors secreted by the plant pathogenic bacterial genus *Xanthomonas*, which causes disease in plants such as rice and cotton (Boch and Bonas, 2010; Bogdanove et al., 2010). TALEs are injected into host cells through the type III secretion system and interfere with cellular activities by activating the transcription of specific target genes (Bogdanove et al., 2010). They have specific structural features, including secretion and translocation signals in the N-terminal region, nuclear localization signals (NLS) and an acidic transcription-activation domain (AD) in the C-terminal region and a central DNA binding domain (DBD) with 33−35 nearly identical long amino acid repeats, followed by the last module which contains only 20 amino acids (refers to as "half repeat") (Fig. 1A). The repeat variable di-residue (RVD) at positions 12 and 13 of each repeat dictates the specificity of repeat binding to a nucleotide in the DNA target (Fig. 1A). This RVD-nucleotide preference was first identified as the DNA sequence recognition code of TALEs in 2009 by two independent groups (Boch et al., 2009; Moscou and Bogdanove, 2009). According to this code, the HD repeat specifies C, NG specifies T, NI specifies A, NN specifies G or A, N* specifies C, IG specifies T, and NS specifies A, C, G, or T. Although many natural RVDs have been discovered, four of them, HD, NN, NI, and NG, account for 75% of the total (Moscou and Bogdanove, 2009) (Fig. 1A).

The DNA binding specificity of a TALE is determined by its repeat number and the sequence of the RVD: the repeat number determines the length of the target sequence, while the RVD corresponds directly to the nucleotide in the target site. Moreover, recognition sites are always preceded by a thymine (T) before the first repeat in the array and this is the only critical rule for TALE targeting (Boch et al., 2009).

The three-dimensional structures of TALE−DNA complexes, reported in two back-to-back papers in 2012 (Deng et al., 2012a; Mak et al., 2012), provided detailed additional information supporting the one-to-one RVD-nucleotide recognition code. It was shown that each TAL repeat comprises two helices connected by a short RVD-containing loop, and all repeats form a right-handed superhelical structure that tracks along the sense strand of the DNA duplex, with the RVDs contacting the major groove. The two hypervariable residues in the RVD loops have different biochemical roles. The second amino acid of the RVD (position 13) mediates specific recognition of the sense strand DNA base, whereas the first amino acid (position 12) does not directly contact the DNA but instead helps to stabilize the repeat structure.

From a target DNA sequence, researchers can in principle analyze the number and order of the required RVDs and design a TALE protein to target a given site in any organism. The simplicity and the modularity of the repeats enable rapid construction of TALE proteins fused to a number of functional domains to achieve site-specific modifications of the genome. These functional domains include repressors, activators, nucleases, nickases, recombinases, methylases, integrases and others (Fig. 1B).

Thus, using HD = C, NN = G, NI = A, and NG = T as code, customizable TALENs have been created and widely used in TALEN-mediated TGM. In addition, since the HD and NN are regarded as strong RVDs, while NI, NG or NK are weak ones, Streubel et al. (2012) suggested that the former should be employed and stretches ($\geq$6) of weak RVDs should be avoided in the design of efficient TALEs. The uses of some rarely used RVDs have also been studied. For example, NK and NH were found to be more specific for G than NN but have lower affinity (Moscou and Bogdanove, 2009; Miller et al., 2011; Zhang et al., 2011; Streubel et al., 2012), while N* or NG were shown to bind 5mC and so overcome methylation of the cytosine (Deng et al., 2012b; Valton et al., 2012).

TALEN ASSEMBLY PLATFORMS

Because of the long and highly repetitive nature of the DBDs of TALEs, it is a major challenge to construct engineered TALENs by ordinary PCR and traditional cloning techniques. Although custom-engineered TALEs have become available commercially through Cellectis Bioresearch and Life