Methods 69 (2014) 76–84

Contents lists available at ScienceDirect

Methods

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

Methods for targeted mutagenesis in zebrafish using TALENs

Woong Y. Hwang [a,b], Randall T. Peterson [a,b,c], Jing-Ruey Joanna Yeh [a,b,*]

[a] Cardiovascular Research Center, Massachusetts General Hospital, 149 13th street, Room 4201, Charlestown, MA 02129, USA
[b] Department of Medicine, Harvard Medical School, 25 Shattuck St, Boston, MA 02115, USA
[c] Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA

ARTICLE INFO

Article history:
Received 4 February 2014
Revised 3 April 2014
Accepted 9 April 2014
Available online 18 April 2014

Keywords:
Site-specific nuclease
Transcription activator-like effector nuclease
Knockout
Genome engineering
Gene targeting
Gene disruption
Zebrafish

ABSTRACT

The transcription activator-like effector (TALE) nucleases, or TALENs, are customizable restriction enzymes that may be used to induce mutations at nearly any investigator-specified DNA sequence in zebrafish. The DNA-binding specificities of TALENs are determined by a protein array comprised of four types of TALE repeats, where each repeat recognizes a different DNA base. Here, we describe methods for constructing TALEN vectors that have been shown to achieve high success rates and mutation efficiencies in zebrafish. In addition, we discuss simple techniques and protocols that can be used to detect TALEN-induced mutations at almost any genomic locus. These methods should enable zebrafish researchers to quickly generate targeted mutations at their genes-of-interest.

1. Introduction

Engineered site-specific endonucleases are essential tools for targeted gene editing in zebrafish, human cells and many other important model organisms (reviewed in [1–4]). These enzymes can be used to create double-stranded DNA breaks at investigator-specified genomic loci, enabling sequence modifications and introduction of exogenous DNA sequences at their target sites [1–4]. To date, three major types of customizable endonucleases have been developed, namely zinc finger nucleases (ZFNs) [5–13], transcription activator-like effector nucleases (TALENs) [10,14–21] and clustered regularly interspaced short palindromic repeats (CRISPR) RNA-guided Cas9 (CRISPR/Cas) nucleases [22–32]. All three classes have been shown to cause efficient target gene disruption in zebrafish [5,6,8,9,14,17,19,27,33–39]. Nevertheless, each platform differs in the ease of their construction methods, potential off-target activities and the theoretical targeting range. In terms of nuclease construction, highly effective ZFNs are generally the most difficult to obtain, while CRISPR/Cas are the easiest. Both TALENs and CRISPR/Cas can reach very high on-target activities. However, in some cases, CRISPR/Cas nucleases have been shown to elicit very high off-target activities [40–43]. Among these approaches, TALENs exhibit the broadest targeting range, with almost no restrictions in target sequence [18]. Thus, TALENs may sometimes be the only possible choice for targeting a particular sequence of interest. Investigators can use the free online software ZiFiT Targeter (http://zifit.partners.org/ZiFiT/) to identify potential target sites for all of these approaches [6,9,19,27,44–46].

TALENs are chimeric proteins comprised of the endonuclease domain of *Fok*I restriction enzyme and a site-specific DNA-binding domain derived from the transcription activator-like effectors (TALEs) of plant pathogens *Xanthomonas* species. The DNA-binding domain of TALEs is composed of a series of TALE repeats where each recognizes a specific DNA nucleotide [47,48]. These 32 to 35-amino acid repeats are almost identical except for two residues at positions 12 and 13 which govern DNA binding specificity [49,50]. We and others have demonstrated that customized TALENs based on this simple code can confer robust on-target activities in zebrafish [14,17,19,33,35,36]. Based on our experience, TALEN-induced mutation rates in somatic cells are found to be as high as 76% [33]. Moreover, TALENs induce heritable mutations in zebrafish efficiently [33]. Thus, TALENs are accessible and powerful research tools for zebrafish genome editing. Here, we discuss potential considerations and our preferred methods for generating zebrafish gene mutations using TALENs.

* Corresponding author.
E-mail address: JYEH1@mgh.harvard.edu (J.-R.J. Yeh).

http://dx.doi.org/10.1016/j.ymeth.2014.04.009
1046-2023/

2. Considerations for choosing a TALEN framework

Currently, there are several public resources for TALEN construction (see review [2] or the website TALengineering.org). TALENs constructed on different platforms may have differences in their architectures, such as the sequences and the numbers of the TALE repeats, the DNA-binding domain outside of the TALE repeats, and modifications in the *Fok*I nuclease domain. Some of these differences may affect the stability of the TALEN vectors, which can be subject to recombination due to the highly repetitive nature of the TALE repeats. Other differences may affect the efficiencies of the enzymes themselves. We have adopted the TALEN framework initially developed by Miller et al. based on the TALE13 of *Xanthomonas axonopodis* [21]. In this scaffold, the DNA-binding domain contains a truncated, 63-amino acid C-terminal segment, which has been found to increase TALEN activities in a serial deletion analysis. Moreover, slight sequence variations have been built into the TALE repeat modules to prevent recombination of TALEN vectors. To date, this TALEN platform has been successfully exploited in numerous model systems such as human cells, zebrafish, rats, and worms [10,16,18–20,33].

3. Considerations for choosing a target site

To construct customized TALENs for a gene of interest, the first step is to identify candidate target sites in its genomic sequence. Researchers can conduct this search using one of the two "TALE Nucleases" functions in ZiFiT Targeter (http://zifit.partners.org/ZiFiT/) depending on the researchers' choice of TALEN construction methods (as discussed below). It is important to note that there are sometimes polymorphisms in a zebrafish colony [33]. Thus, investigators may wish to check their stock to confirm that there is no polymorphic sequence in the planned TALEN target sites. In addition, it has been shown that TALE DNA-binding domains are sensitive to DNA methylation [51]. Methylation on the cytosine of a CpG dinucleotide can be identified *via* bisulfite sequencing. Alternatively, researchers may want to avoid target sites containing CpG sequences if there are other options [36]. Beyond avoidance of polymorphisms and methylation sites, it is important to avoid highly repetitive sequences, which may increase risk of off-target cleavage. However, some earlier guidelines proposed by Cermak et al. based on a computational analysis of natural TAL effectors have been found to be non-critical [18,52].

4. Construction of customized TALEN vectors

Since the *Fok*I nuclease domain cleaves DNA in a dimeric form, two customized TALENs are needed for any given target locus. The DNA-binding domains of these two TALENs should bind to the left and the right target half sites, respectively. This will facilitate the dimerization of the nuclease domains in the spacer region between the two target half sites (Fig. 1). Thus, an optimal spacer length is important for efficient DNA cleavage activity. Previously Miller et al. have shown that, using their TALEN framework, the nucleases are the most active when the spacer lengths are between 14 and 19 basepairs (bps) [21]. We have found that spacer lengths ranging between 16 and 18 bps can be effective in zebrafish [19]. The same spacer range has also been validated in a large-scale test of TALENs that target endogenous human genes in cultured cells [18], and is used as the default setting in ZiFiT Targeter.

The number of TALE repeats in the DNA-binding domain can also influence the efficiency of the TALENs constructed [18]. Native TALEs often contain 17.5 TALE repeats where the final "0.5" repeat is only 20 amino acids long [21]. In addition, the first repeat typically binds to a thymidine nucleotide [21]. Thus, engineered TALENs normally have a fixed first repeat, which is followed by a customized TALE repeat array (Fig. 2). By evaluating the mutation efficiencies and cytotoxicities induced by TALENs harboring 8.5–19.5 TALE repeat arrays, Reyon et al. have proposed that TALENs composed of 14.5- to 16.5-repeat arrays (which will recognize a total of 16–18 DNA nucleotides including the 5′ T) are most likely to be effective [18]. Following this guideline, our overall success rate of customized TALEN pairs that can mutate their target loci in zebrafish is above 80% ([19,33] and unpublished results).

Construction of customized TALENs involves the generation of a TALE repeat array and the cloning of this array into a TALEN expression vector that has the coding sequences for the rest of the TALE DNA-binding domain and the *Fok*I nuclease domain. We use TALE repeat arrays generated using three different methods developed by Joung and colleagues. The first method is a restriction enzyme and ligation-based protocol called REAL [19,44]. This method has the lowest throughput, but it does not require a large number of plasmid archives as the starting materials. Thus, it is suitable for researchers who are interested in producing only a few TALEN pairs. The second method, called REAL-Fast, is a modified version of the REAL method [44]. The major difference between these two methods is that REAL-Fast takes advantage of a larger number of plasmid archives as the starting materials (32 plasmids for REAL and 380 plasmids for REAL-Fast). Thus, REAL-Fast will not only increase the throughput but also shorten the construction time to 7 days as opposed to 9 days using REAL for arrays containing 12.5–16.5 TALE repeats. The third method is called Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) assembly [18,33,45]. The FLASH method can be conducted in a 96-well plate by automation or manually. Since the assembly is performed on solid phase, it does not call for cloning of the intermediate products as the other two methods do. Using FLASH, 96 arrays containing 16.5 TALE repeats or less can be built in one day. It is worth noting that these methods require either none or at most one PCR amplification throughout the whole process, thus lowering the chances of sequence errors introduced by PCR. Detailed protocols for constructing customized TALE repeat arrays and TALEN vectors using these methods have been described elsewhere [44,45]. The Joung lab REAL assembly TALEN kit is available from Addgene (http://www.addgene.org/talengineering/TALENkit/). Plasmid archives required for the REAL-Fast and the FLASH methods can be requested from the Joung lab (http://eGenome.org).

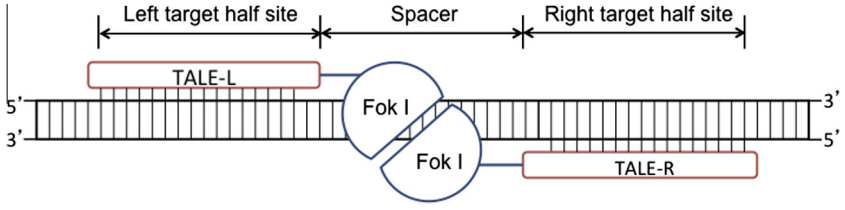

Fig. 1. A schematic representation of TALENs binding to their cognate target sites. The DNA-binding domains of two TALENs (TALE-L and TALE-R) bind to the two target half sites, facilitating dimerization of their *Fok*I nuclease domains. The dimerized nucleases can then create double-stranded DNA breaks in the spacer region.