



Published in final edited form as:

*Curr Protoc Mol Biol.* 2012 October ; 0 12: Unit–12.15. doi:10.1002/0471142727.mb1215s100.

## Current Protocols in Molecular Biology Engineering Designer Transcription Activator-Like Effector Nucleases (TALENs)

Deepak Reyon<sup>1,2</sup>, Cyd Khayter<sup>1</sup>, Maureen R. Regan<sup>1</sup>, J. Keith Joung<sup>1,2,\*</sup>, and Jeffrey D. Sander<sup>1,2,\*</sup>

<sup>1</sup>Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129 USA

<sup>2</sup>Department of Pathology, Harvard Medical School, Boston, MA 02115 USA

### Abstract

Engineered transcription activator-like effector nucleases (TALENs) are broadly useful tools for performing targeted genome editing in a wide variety of organisms and cell types including plants, zebrafish, *C. elegans*, rat, human somatic cells, and human pluripotent stem cells. Here we describe a detailed protocol for the serial, hierarchical assembly of TALENs that requires neither PCR nor specialized multi-fragment ligations and that can be implemented by any laboratory. This restriction enzyme and ligation (REAL) protocol can be practiced using a small plasmid library and user-friendly, web-based software that both identifies target sites in sequences of interest and generates printable graphical guides that facilitate assembly of TALENs. All plasmids required to perform the REAL assembly method are publicly available through the Addgene plasmid distribution service (<http://www.addgene.org/talengineering>). Alternatively, to decrease the time and effort required, users can practice REAL using a library of plasmids encoding pre-assembled TALE repeats. With the platform of reagents, protocols, and software we describe, researchers can easily engineer multiple TALENs within two weeks or less using standard cloning techniques.

### Keywords

TALEN; TALENs; engineered TAL nucleases; engineered TALE nucleases; REAL; REAL-Fast; protein engineering; DNA-binding domains; FLASH

### Introduction

Engineered transcription activator-like effector nucleases (TALENs) have recently generated much interest as a broadly applicable technology for highly efficient genome editing (Baker, 2012; DeFrancesco, 2011). TALENs, like zinc finger nucleases (Chapter 12.3), are customizable restriction enzymes consisting of an engineered DNA-binding

\*Correspondence to: JJOUNG@PARTNERS.ORG or JSANDER@PARTNERS.ORG.

#### Author Contributions

J.D.S. and J.K.J. conceived of the assembly strategy and designed the plasmid DNAs. J.D.S., D.R., M.R.R. and C.K. experimentally validated and refined the strategy. J.D.S., D.R., and J.K.J. designed the ZiFiT Targeter software and wrote the paper. We thank Morgan Maeder for helpful discussions.

#### Competing Financial Interests

J.D.S. and J.K.J. are inventors on a patent application describing the related FLASH TALE repeat array assembly method.

#### Internet Resources

<http://zifit.partners.org>

Provides access to the ZiFiT software program for engineering TALENs

domain fused to a non-specific nuclease domain (Figure 1A) (Christian et al., 2010; Li et al., 2011a; Mahfouz et al., 2011; Miller et al., 2011; Mussolino et al., 2011). Site-specific double-stranded DNA breaks induced by TALENs have been used to introduce sequence alterations at investigator-specified endogenous genes in a variety of organisms and cell types including yeast (Li et al., 2011b), plants (Cermak et al., 2011), *C. elegans* (Wood et al., 2011), zebrafish (Huang et al., 2011; Sander et al., 2011), rats (Tesson et al., 2011), and human somatic (Cermak et al., 2011; Miller et al., 2011; Mussolino et al., 2011; Reyon et al., 2012) and pluripotent stem cells (Hockemeyer et al., 2011).

The DNA-binding domain of a TALEN consists of an array of repeat sequences derived from naturally occurring transcription activator-like effectors (TALEs) (Bogdanove et al., 2010; Scholze and Boch, 2011). These highly conserved 33–35 amino acid TALE repeat domains each bind to a single DNA base with binding specificity associated with the identity of two non-conserved amino acid at positions 10 and 11 (Boch et al., 2009; Moscou and Bogdanove, 2009). DNA-binding domains with novel customized specificities can be engineered by joining TALE repeats into more extended arrays (Bogdanove and Voytas, 2011) (Figure 1B). One challenge for researchers interested in utilizing TALENs is the need to construct plasmids encoding long arrays of TALE repeats that are each highly similar in sequence.

We recently described a rapid and simple method for assembling DNAs encoding extended TALE repeat arrays (Sander et al., 2011). Our method uses a serial, hierarchical ligation strategy in which DNA fragments encoding single TALE repeats are joined together using standard restriction digest and ligation techniques that can be practiced simply and inexpensively by any laboratory, an approach that we now refer to as the REAL (Restriction Enzyme And Ligation) method. We have also recently developed a more rapid and less labor-intensive version of REAL that we refer to as REAL-Fast. In contrast to REAL, which uses plasmids encoding single TALE repeats, REAL-Fast uses an archive of plasmids encoding pre-assembled multiple TALE repeats. (This archive is the same one used to practice our recently described Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) method for assembling engineered TALE repeat arrays (Reyon et al., 2012).) To facilitate the use of REAL and REAL-Fast, we have also developed web-based software that aids not only in identification of potential target sites within a sequence of interest but also produces a customized graphical guide illustrating the series of ligation steps required to construct each TALE repeat array by either REAL or REAL-Fast.

## Strategic Planning

A variety of different methods for constructing DNA sequences encoding TALE repeat arrays have been described (Cermak et al., 2011; Geissler et al., 2011; Huang et al., 2011; Li et al., 2011a; Miller et al., 2011; Morbitzer et al., 2011; Sun et al., 2012; Weber et al., 2011; Zhang et al., 2011). However, in choosing which of these methods to use, a potential user should carefully consider not only the technical details of the assembly strategy but also the architecture and amino acid sequence of the final assembled TALE repeat arrays produced by the specific approach. We provide a detailed overview of both of these issues here.

Nearly all of the methods for TALE repeat array assembly described to date rely on the use of a specialized multi-fragment ligation strategy known as Golden Gate cloning. With these approaches, DNA encoding subsets of TALE repeats (ranging in length from four- to ten-mers) are initially assembled in parallel using multi-fragment ligation reactions and then subsequently joined together to create the final desired plasmid. Many of the methods are optimized for construction of arrays with either a fixed number or a fixed multiple of repeats, limiting their flexibility to construct arrays of any desired length. Only one of these

methods has associated publicly available software for identifying potential TALEN target sites in a sequence of interest (Cermak et al., 2011). However, this software does not provide user-friendly guidance on how to perform the complex multi-fragment ligation process required to assemble TALENs.

Another important consideration in choosing a platform for assembly is the architecture of the TALE-based DNA-binding domain one uses to construct TALENs. A multitude of different engineered TALE architectures have been described and utilized in the literature with two sources of variability among these different frameworks: (1) the specific amino acid sequences of TALE repeats used within assembled arrays can differ at certain less-stringently conserved positions within each domain and (2) the length and sequences of additional amino-terminal and carboxy-terminal TALE domains that flank the TALE repeat array and that are important for DNA-binding activity (Christian et al., 2010; Miller et al., 2011; Mussolino et al., 2011; Zhang et al., 2011). The choice of architecture is important and can influence the activities of TALENs (Miller et al., 2011; Mussolino et al., 2011). We note that our REAL and REAL-Fast platforms (and also our recently described FLASH assembly platform), in contrast to all other publicly available methods, utilize an architecture of specific TALE repeats and amino-terminal and carboxy-terminal TALE-derived sequences that has been used successfully to construct TALENs with high activities in *C. elegans* (Wood et al., 2011), zebrafish (Sander et al., 2011), rats (Tesson et al., 2011), and human somatic (Miller et al., 2011; Reyon et al., 2012) and pluripotent stem cells (Hockemeyer et al., 2011).

Here we describe a detailed protocol for identifying potential TALEN targets within a sequence of interest and for assembling TALENs to those sites using our REAL or REAL-Fast methods. Both methods use standard restriction digest and ligation reactions and therefore do not require specialized expertise or the need to perform multi-fragment ligation reactions such as Golden Gate. Our web-based software program (freely available without registration) identifies TALEN target sites in user-defined sequences and generates printable, color-coded graphical guides that provide a roadmap for assembly of desired TALENs by either REAL or REAL-Fast. These customized guides also give the names of specific plasmids required to assemble each particular TALEN. All plasmids required to practice REAL are publicly available to academic researchers from non-profit plasmid distribution service Addgene (<http://www.addgene.org/talengineering>). The archive of plasmids encoding pre-assembled TALE repeats that is required to practice REAL-Fast is also available by request from the Joung lab (<http://www.TALEngineering.org>).

## Basic Protocol: Engineering Customized TALENs

Both the REAL and REAL-Fast protocols can be conceptually divided into three steps: (1) identification of potential TALEN targets within the sequence of interest using our web-based ZiFiT Targeter software; (2) assembly of plasmids encoding TALE repeat arrays following graphical roadmaps generated by ZiFiT Targeter; and (3) cloning of DNA fragments encoding TALE repeat arrays into a TALEN expression vector.

### Identifying TALEN target sites using the ZiFiT Targeter program

We have updated our previously described Zinc Finger Targeter (ZiFiT) program to also enable identification of TALEN target sites. To reflect this change, we have re-named our program Zinc Finger & TALE Targeter (ZiFiT Targeter). ZiFiT Targeter can be used without requirement to register and is available at <http://zifit.partners.org>. Because TALENs can be targeted to a broad range of potential sites, ZiFiT Targeter restricts its output by default to five highest-ranked potential sites. These rankings are based on published data describing the optimal length of arrays and the length of spacer sequences between the

TALEN binding sites and with the goal of avoiding substantially overlapping, and therefore similar, sites (Miller et al., 2011; Reyon et al., 2012). ZiFiT Targeter also provides users with customized graphical guides for assembly of each TALEN.

### Assembly of plasmid DNA encoding TALE repeat arrays

Plasmid DNAs encoding TALE repeat arrays are rapidly assembled using a serial, hierarchical assembly strategy based on simple restriction enzyme digests and standard ligations (Figure 2). The details of which particular plasmids to use for these reactions and the specific order in which to perform the ligations are provided by graphical guides generated by ZiFiT Targeter for each TALEN. Completion of the procedure illustrated by these guides will yield a plasmid encoding the final desired TALE repeat array flanked by unique restriction sites that can be used for cloning into a TALEN expression vector.

### Cloning of DNA encoding TALE repeat arrays into a TALEN expression vector

In the final step, a DNA fragment encoding the assembled TALE repeat array is cloned into a TALEN expression vector and then verified by DNA sequencing. Our TALEN expression vectors provide the final carboxy-terminal “0.5” TALE repeat domain, additional amino-terminal and carboxy-terminal TALE-derived sequences required for optimal DNA binding, and the wild-type FokI nuclease domain. TALENs expressed from these vectors also possess a triple FLAG epitope tag and a nuclear localization signal, both encoded at the amino-terminus. A TALEN encoded by these vectors can be expressed in many cell-types using a CMV promoter that is present on the vector. Alternatively, the TALEN coding sequence can be transcribed into RNA *in vitro* using a T7 promoter also encoded on the expression plasmid.

## MATERIALS

- Plasmids encoding individual TALE repeats (available through Addgene: <http://www.addgene.org/talengineering>)
- Archive of preassembled TALE arrays for practicing REAL-Fast (available through the Joung lab; see <http://www.TALEngineering.org>)
- TALEN Expression Vectors (available through Addgene: <http://www.addgene.org/talengineering>)
- Chemically competent bacterial strain XL-1 Blue (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacIq lacZDM15 Tn10* (Tet<sup>R</sup>)]; Stratagene cat no. 200249)
- Carbenicillin (Sigma, cat. No C1389)
- Restriction enzymes from New England Biolabs: BamHI (cat. no. R0136S), BbsI (cat. no. R0539L), BsaI (cat. no. R0535L), BsmBI (cat. no. R0580L), and KpnI-HF (cat. no. R3142L)
- Bovine Serum Albumin (10 mg/ml; included with enzymes from New England Biolabs)
- NEBuffer2, NEBuffer 3, and NEBuffer 4 (10× restriction enzyme buffers included with enzymes from New England Biolabs)
- Quick Ligation Kit (New England Biolabs, cat. No. M2200L)
- QIAprep Spin Miniprep Kit (Qiagen, cat. No. 27106)
- LB medium powder (Difco, cat No. 244620)

- LB agar medium powder (Difco, cat No 244520)
- AccuGel 29:1 acrylamide:bis-acrylamide solution (National Diagnostics, cat no. EC-852)
- 10% ammonium persulfate (Fisher cat. No 7727-54-0)
- TEMED (Fisher cat. No. BP150-100)
- 100% ethanol (Pharmco, cat. No. 111ACS200)
- 70% ethanol
- Sequencing primer OK163: 5' CGCCAGGGTTTTCCCAGTCACGAC 3'
- Sequencing primer JDS2978: 5' TTGAGGCGCTGCTGACTG 3'
- Sequencing primer JDS2980: 5' TTAATTCAATATATTCATGAGGCAC 3'
- Sequencing primer JDS2778: 5' CTGGCGCAATGCGCTCAC 3'
- Sequencing primer JDS2979: 5' AAGCAATGGCGACCACCTGTTC 3'
- 96-well PCR thermocycler
- Orbital platform shaker with adjustable speed
- Sterile bacterial culture tubes
- 1.5ml Eppendorf Tubes
- Tabletop Centrifuge

### Identifying TALEN target sites

- 1) Identify any repeat sequences within the genomic sequence of interest by entering it into the RepeatMasker Web Server (<http://www.repeatmasker.org/>). Repeat sequences should be excluded from any sequence that is to be analyzed for potential TALEN target sequences by ZiFiT Targeter.

Note that repetitive sequences are not identified if the sequence submitted is shorter in length than the repetitive element. To avoid this problem, we recommend that you submit your target sequence with a few hundred base pairs of additional flanking genome sequence.

Make sure that you submit genomic sequence (and not cDNA) sequence to ensure that TALEN targets identified by ZiFiT Targeter are present in the genome.

We strongly recommend that you sequence your genomic target region from the organism or cell type you plan to perform experiments in to avoid targeting sites that have polymorphisms that may therefore differ from published genome sequences.

- 2) Visit the ZiFiT Targeter website at <http://zifit.partners.org>.
- 3) Click on the ZiFiT option on the top menu and then click on the “Design TALE Nucleases” option under the “TALE Assembly” menu.
- 4) Paste one to twelve nucleotide sequences in FASTA format into the text box labeled ‘Sequence’. All numbers and characters in the sequence that are not G, A, T, or C will be ignored. The user must indicate the nucleotide position at which they wish the break to occur by framing it with brackets (e.g.--[A] or

[G]). ZiFiT Targeter will attempt to identify sites that place this nucleotide within the spacer sequence between the TALEN target half-sites.

- 5) Ensure the option “Mask redundant sites” is checked to make sure that the sites identified by ZiFiT Targeter are different from each other by at least 6 base pairs (3 base pairs in each TALEN monomer binding site).
- 6) By default, for each target site entered, ZiFiT Targeter will return five potential targets using criteria shown to work robustly in Reyon et al (Reyon et al., 2012). However, a user can display all potential target sites identified by the program by checking the box for “Relax Constraints”. With this option checked, ZiFiT Targeter will return all potential TALEN target sites that consist of half-sites of lengths 12–22 base pairs (with the conserved 5' T nucleotide included as part of the half-site) and spacer sequences of lengths 12–23 base pairs.
- 7) Click the Submit button. The output reports the top five TALEN targets for each of the sites in different tabs below the sequence entry box. To help users visualize the target sites within the context of the query sequence, the target sites are aligned to the query sequence and the position of the desired double stranded break is indicated with an asterisk (Figure 3).
- 8) Click on the target site to obtain the DNA sequence and a customized graphical guide for assembling a particular TALEN (Figure 4). The guide and sequence will open in a new window within which users can choose between either REAL or REAL-Fast assembly using the buttons on the top of the page. The numbers above each TALE repeat unit in the first row of the graphical guide identify the source plasmids. If the “REAL” button is selected these numbers refer to the REAL plasmid set available from Addgene. Alternatively if “REAL-Fast” is selected the identification number refers to the FLASH plasmid archive (available from the Joung lab).

*A survey of 144 TALENs pairs constructed using this architecture demonstrated robust mutagenesis across a broad range of parameters independent of target composition with the exception of a 5' T (Reyon et al., 2012). ZiFiT identifies targets with parameters that were demonstrated to work robustly in this large scale study.*

### Construction of TALE Arrays

- 9) Using the graphical output provided by ZiFiT Targeter as a guide, perform the series of ligations indicated in the first row of the figure by using steps 10 – 19 below. Note that the numbers above each TALE repeat unit/array in the first row of the graphical guide identify the names of the plasmids harboring the TALE repeats. For example, in the graphical output shown in Figure 4A, one would perform seven ligations of the following pairs of plasmids: 7 and 14, 16 and 22, 30 and 11, 17 and 24, 26 and 15, 16 and 24, and 30 and 12). For each pair to be ligated, the TALE repeat on the left is the amino-terminal repeat and the one on the right is the carboxy-terminal repeat. In each ligation, a DNA vector backbone encoding the amino-terminal repeat is ligated to a DNA fragment encoding the carboxy-terminal repeat.
- 10) Digest plasmid(s) encoding the amino-terminal TALE repeat(s) with BamHI and BsaI using the conditions listed below. Incubate the reaction for 2 hours at 37°C.

<u>Component</u>	<u>Amount</u>
Plasmid encoding amino-terminal TALE repeat(s)	1.5 µg
NEB Buffer 2	4 µl
Bovine Serum Albumin (10mg/ml)	4 µl
BsaI (5 U/µl)	2 µl
BamHI (20 U/µl)	2 µl
Add nuclease-free water to a total volume of	40 µl

- 11) Isolate the vector backbone from the restriction digest of step 10 on a 5% non-denaturing polyacrylamide gel and purify the DNA according to the protocol found in Box 2 of Maeder et al (Maeder et al., 2009). Final DNA pellets should be resuspended in 20 µl of nuclease-free water.
- 12) Digest plasmid(s) encoding the carboxy-terminal TALE repeat(s) with BbsI and BamHI using the conditions listed below. Incubate the reaction for 2 hours at 37°C.

<u>Component</u>	<u>Amount</u>
Plasmid encoding carboxy-terminal TALE repeat	1.5 µg
10X Buffer (NEBuffer 4)	4 µl
Bovine Serum Albumin (10mg/ml)	4 µl
BbsI (5 U/µl)	2 µl
BamHI (10 U/µl)	2 µl
Add nuclease-free water to a total volume of	40 µl

- 13) Isolate the DNA fragment encoding the TALE repeat(s) from the restriction digest of step 12 on a 5% non-denaturing polyacrylamide gel and purify the DNA using the protocol found in Box 2 of Maeder et al. (Maeder et al., 2009). Final DNA pellets should be resuspended in 20 µl of nuclease-free water. Note that the size of the DNA fragment will be equal to  $N \times \sim 102$  bps where N is the number of TALE repeats encoded on the fragment.
- 14) Ligate the purified DNA fragment isolated in step 13 into the purified vector backbone isolated in step 11 as tabulated below. Also perform a control ligation using vector backbone alone (i.e.--without fragment). Allow the ligation reactions to incubate for 15 minutes at room temperature.

<u>Component</u>	<u>Amount</u>
Purified vector backbone (from step 11)	1 µl
Purified fragment (from step 13) or water	3 µl
Quick Ligase Buffer (NEB)	4.5 µl
T4 DNA Ligase (400U/ul)	0.5 µl
Total	9 µl

- 15) Transform each ligation from step 14 into 90 µl chemically competent XL1-Blue cells. Mix ligations with competent cells and leave on ice for 5 minutes. Perform

heat shock at 42°C for 1 minute then return transformations to ice for 1 minute. Add 500 µl LB and recover with agitation for 45 minutes at 37°C. Plate 200 µl of each transformation on an LB agar plate supplemented with 100 µg ml<sup>-1</sup> carbenicillin and incubate overnight at 37°C for 12 – 16 hours.

- 16)** If the actual ligation/transformation of step 15 yields at least 10-fold more colonies than the control ligation, inoculate two single colonies from the actual ligation/transformation plate into 4 ml of LB supplemented with carbenicillin 100 mg ml<sup>-1</sup> and grow overnight with agitation at 37°C.

To reduce the risk of plasmid deletions, do not allow the cultures to grow for more than 14 hours.

- 17)** Isolate plasmid DNA from overnight cultures using a QIAprep Spin Miniprep Kit and following the manufacturer's instructions.
- 18)** To determine whether the plasmids isolated in step 17 have successfully taken up the fragment encoding the carboxy-terminal TALE repeat(s), digest the candidate plasmids with XbaI and BamHI as detailed below and incubate at 37°C for 1 hour.

<u>Component</u>	<u>Amount</u>
Plasmid	1 µg
10X Buffer (NEB Buffer 4)	4 µl
Bovine Serum Albumin (10mg/ml)	4 µl
XbaI (20 U/µl)	2 µl
BamHI (20 U/µl)	2 µl
Add nuclease-free water to a total volume of	40 µl

- 19)** Visualize the products of the restriction digests from step 18 on a 5% non-denaturing polyacrylamide gel. Plasmids that have successfully taken up the fragment encoding the carboxy-terminal TALE repeat(s) should yield a  $[(M + N) \times 102] + 33$  bp fragment (where M and N are the numbers of TALE repeats encoded by the vector backbone and fragment, respectively, used for the ligation).
- 20)** Following completion of the first set of ligations, perform the series of ligations indicated in the second row of the graphical guide by using steps 10 – 19 above.
- 21)** Continue performing ligations in subsequent rows of the graphical guide using steps 10 – 19 above until the final assembled array is completed.

To ensure that repeats have been assembled correctly, intermediate constructs can be sequenced with primer OK163.

### Cloning TALE arrays into the nuclease backbone

- 22)** Digest the specific TALEN expression vector indicated at the bottom of the graphical output (pJDS70, pJDS71, pJDS74 or pJDS78) with BsmBI restriction enzyme as detailed below. Incubate the reaction at 55°C for 3 hours.

<u>Component</u>	<u>Amount</u>
------------------	---------------



TALEN Expression Vector	2µg
10× Buffer (NEB Buffer #3)	5 µl
BsmBI (10U/ul)	5 µl
Add nuclease-free water to a total volume of	50 µl

- 23) Isolate the vector backbone from the restriction digest of step 22 on a 5% non-denaturing polyacrylamide gel and purify the DNA using the protocol found in Box 2 of Maeder et al. (Maeder et al., 2009). Final DNA pellets should be resuspended in 20 µl of nuclease-free water.
- 24) Digest plasmid(s) encoding the final assembled TALE repeat array (from step 21 above) with BbsI and BsaI as tabulated below. Incubate the reaction for 2 hours at 37°C.

Component	Amount
Plasmid encoding assembled TALE repeat array	1.5 µg
10X Buffer (NEB Buffer 2)	4 µl
Bovine Serum Albumin (10mg/ml)	4 µl
<i>BbsI</i> (5 U/µl)	2 µl
<i>BsaI</i> (10 U/µl)	2 µl
Add nuclease-free water to a total volume of	40 µl

- 25) Isolate the DNA fragment encoding the TALE repeat array(s) from the restriction digest of step 24 on a 5% non-denaturing polyacrylamide gel and purify the DNA using the protocol found in Box 2 of Maeder et al. (Maeder et al., 2009). Final DNA pellets should be resuspended in 20 µl of nuclease-free water. Note that the size of the DNA fragment will be equal to  $N \times \sim 102$  bps where N is the number of TALE repeats encoded on the fragment.
- 26) Ligate the purified DNA fragment isolated in step 25 into the purified TALEN expression vector backbone isolated in step 23 as tabulated below. Also perform a control ligation using vector backbone alone (i.e.--without fragment). Allow the ligation reactions to incubate for 15 minutes at room temperature.

Component	Amount
Purified vector backbone (from step 23)	1 µl
Purified fragment (from step 25) or water	3 µl
Quick Ligase Buffer (NEB)	4.5 µl
T4 DNA Ligase (400U/ul)	0.5 µl
Total	9 µl

- 27) Transform each ligation from step 26 into 90 µl chemically competent XL1-Blue cells. Mix ligations with competent cells and leave on ice for 5 minutes. Perform heat shock at 42°C for 1 minute then return transformations to ice for 1 minute. Add 500 µl LB and recover with agitation for 45 minutes at 37°C. Plate 200 µl of each transformation on an LB plate supplemented with 100 µg ml<sup>-1</sup> carbenicillin and incubate overnight at 37°C for 12 – 16 hours.

- 28) If the actual ligation/transformation of step 27 yields at least 10-fold more colonies than the control ligation, inoculate two single colonies from the actual ligation/transformation plate into 4 ml of LB supplemented with carbenicillin  $100 \text{ mg ml}^{-1}$  and grow overnight with agitation at  $37^\circ\text{C}$ .
- To reduce the risk of plasmid deletions, do not allow the cultures to grow for more than 14 hours.
- 29) Isolate plasmid DNA from overnight cultures using a QIAprep Spin Miniprep Kit and following the manufacturer's instructions.
- 30) To determine whether the TALEN expression plasmids isolated in step 29 have successfully taken up the fragment encoding the TALE repeat array(s), digest the candidate plasmids with KpnI and BamHI as detailed below and incubate at  $37^\circ\text{C}$  for 2 hours.

<u>Component</u>	<u>Amount</u>
Plasmid	0.5 $\mu\text{g}$
10X Buffer (NEBuffer 4)	5 $\mu\text{l}$
Bovine Serum Albumin (10 mg/ml)	5 $\mu\text{l}$
KpnI-HF (20 U/ $\mu\text{l}$ )	1 $\mu\text{l}$
BamHI (20 U/ $\mu\text{l}$ )	1 $\mu\text{l}$
Add nuclease-free water to a total volume of	50 $\mu\text{l}$

- 31) Visualize the products of the restriction digests from step 30 on a 5% non-denaturing polyacrylamide gel. Plasmids that have successfully taken up the fragment encoding the TALE repeat array should yield a  $(650 + [N \times \sim 102])$  bp fragment (where N is the number of TALE repeats encoded in the array.).
- 32) Optional: TALEN expression plasmids can be sequence-verified by DNA sequencing using forward primer JDS2978 and reverse primer JDS2980. This step is not required because no PCR is performed during the assembly process. However, sequencing can be useful to confirm that the correct TALE repeat arrays have been ligated together in the extended array.

*The vector can now be transfected into any mammalian cell type of choice in which the CMV promoter is known to be active (Maeder et al., 2008; Reyon et al., 2012)(Maeder et al. 2008). Alternatively, the TALEN coding sequence can be transcribed into RNA from the T7 promoter as previously described (Foley et. al. 2009)(Sander et al. 2011).*

## Commentary

**Background Information**—TALENs assembled using REAL and REAL-Fast have the same amino acid sequence and architecture as those made using the high-throughput FLASH assembly method(Reyon et al., 2012). Previous work has shown that TALENs made on this architecture can function efficiently in a variety of different organisms and cell types. Reyon et al. recently generated TALENs on this architecture targeted to 48 diverse sites in the EGFP reporter gene. All 48 of these pairs were shown to efficiently disrupt the coding sequence of an integrated EGFP reporter gene in human U2OS cells. Using this same framework, they generated TALEN pairs to the N-terminal region of 96 endogenous genes known to be important in cancer and epigenetic regulation. These TALENs induced high

rates of NHEJ-mediated mutagenesis (average of 22.5%) at 84 of the 96 endogenous gene targets in cultured human cells (Reyon et al., 2012). In addition to demonstrating the robustness of this framework, Reyon et al. conservatively estimated that three TALENs pairs could be assembled for every base pair of random DNA sequence.

### Critical Parameters and Troubleshooting

It is critical to submit genomic sequence (not cDNA) to ZiFiT. Sequences based on cDNA may contain junction sequences that are formed after intron splicing and that are not present in the genomic sequence.

Because TALENs comprise a large number of highly similar repeats, plasmids encoding TALENs are prone to recombination when passaged through *E. coli*. To minimize the chance of recombination, it is critical to propagate plasmids in a strain of *E. coli* that is *recA*-deficient (we use XL-1 Blue cells) and to limit the growth time to no more than 14 hours at 37° C to minimize the time that the cells spend in stationary phase.

Because TALE nuclease monomers can harbor as many as 20 or more TALE repeats, with each repeat encoded by ~102 bp, standard sequencing methods are typically inadequate to verify the full-length construct. The following suggestions can help to enable the sequence verification of TALEN constructs 1) Use a sequencing facility that offers longer sequence reads (>=1000bp) 2) Use forward primer JDS2778 and reverse primer 2979 to prime closer to the repeat region. Note: Because these primer sites are extremely close to the RVD regions, sequencing primers JDS2978 and JDS2980 which prime approximately 100bp and 200bp external to JDS2778 and 2979 are still required to verify the 0.5 domain and cloning junction for the TALEN plasmid..

### Anticipated Results

We have not encountered significant difficulties with assembling any of the various TALEN expression plasmids using the approach described above. Occasionally the vector backbone will be missing the fragment insert. However, this is rare and typically only requires picking a second colony. A factor in ensuring success is to check by restriction digest analysis that the ligations work successfully at each step because one failed ligation reaction can prevent successful assembly of the final desired array. In addition it is critical to sequence the final TALEN expression plasmid to verify that the correct TALE units have been used at each step and that the array was cloned into the correct TALEN expression vector. If the correct units were used we have a 100% success rate engineering TALENs using this protocol.

In a recent report using TALENs, constructed on the same architecture as those made using REAL or REAL-Fast, we were able to robustly disrupt EGFP activity of a single copy integrated reporter in human U2OS cells for 100% of TALEN pairs (48/48). We were also able to achieve high levels of gene disruption for 84 of 96 TALEN pairs targeted to genes involved in human cancer and epigenetic regulation (Reyon et al., 2012). TALENs appear capable of targeting virtually any sequence, provided each half site is preceded by a 5' T.

### Timing Considerations

Plasmids encoding 12.5 to 16.5 TALE repeats can be assembled in ~9 days for REAL and ~7 days REAL-Fast. Longer arrays require an additional two days. The process of cloning the fragment encoding the assembled TALE repeat array into the TALEN expression vector requires 3 days. However, because this final cloning step can be started the same day that the assembly of the TALE repeat array is completed, the assembly of a TALEN expression plasmid containing up to 16.5 TALE repeats can be completed in 11 days using REAL and 9 days using REAL-Fast.

## Acknowledgments

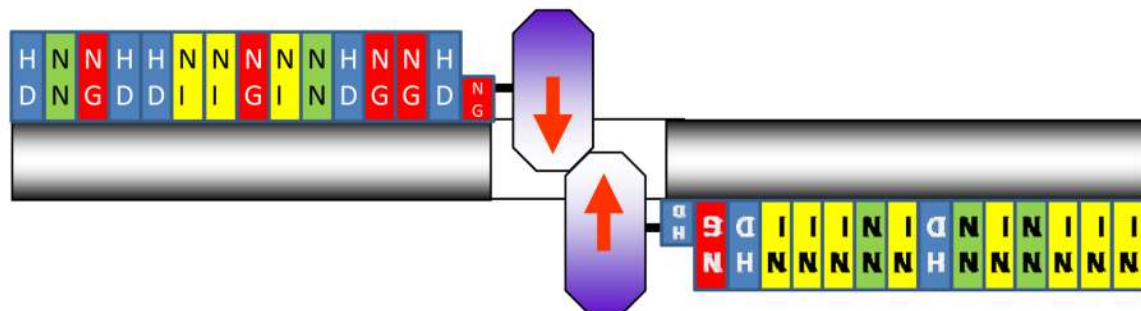
This work was supported by a National Institutes of Health (NIH) Director's Pioneer Award DP1 OD006862 (J.K.J.), NIH R01 GM088040 (J.K.J.), the Jim and Ann Orr MGH Research Scholar Award (J.K.J.), NIH T32 CA009216 (J.D.S.), and National Science Foundation DBI-0923827 (D.R. and J.K.J.).

## REFERENCES

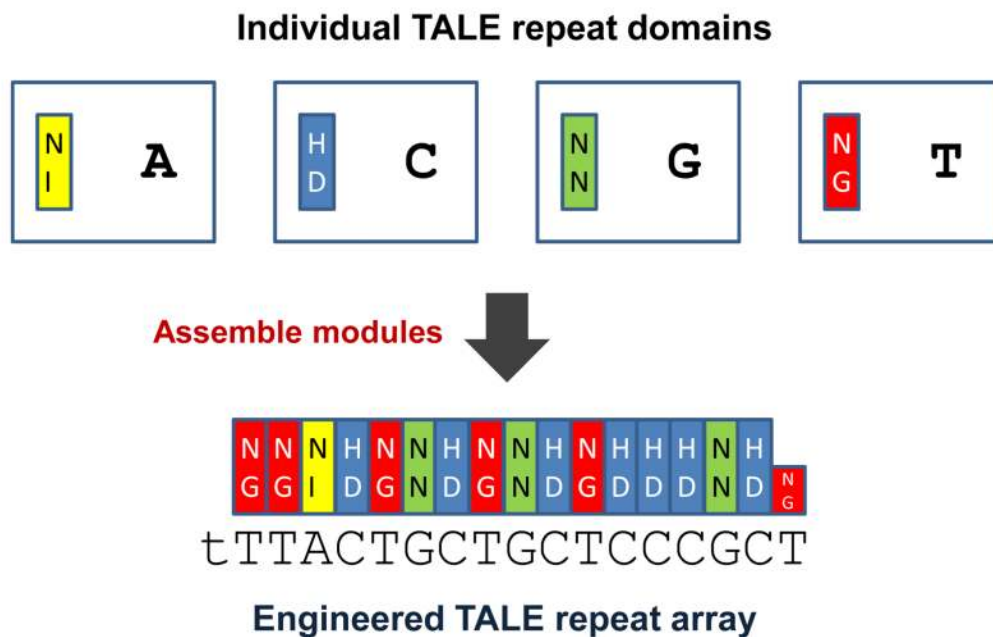
- Baker M. Gene-editing nucleases. *Nature methods*. 2012; 9:23–26. [PubMed: 22312637]
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 2009; 326:1509–1512. [PubMed: 19933107]
- Bogdanove AJ, Schornack S, Lahaye T. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol*. 2010; 13:394–401. [PubMed: 20570209]
- Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science*. 2011; 333:1843–1846. [PubMed: 21960622]
- Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res*. 2011; 39:e82. [PubMed: 21493687]
- Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*. 2010; 186:757–761. [PubMed: 20660643]
- DeFrancesco L. Move over ZFNs. *Nat Biotechnol*. 2011; 29:681–684. [PubMed: 21822235]
- Geissler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens SE, Boch J. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS ONE*. 2011; 6:e19509. [PubMed: 21625585]
- Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*. 2011; 29:731–734. [PubMed: 21738127]
- Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B. Heritable gene targeting in zebrafish using customized TALENs. *Nature biotechnology*. 2011; 29:699–700.
- Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, Yang B. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res*. 2011a; 39:359–372. [PubMed: 20699274]
- Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*. 2011b; 39:6315–6325. [PubMed: 21459844]
- Maeder ML, Thibodeau-Beganny S, Osiaik A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell*. 2008; 31:294–301. [PubMed: 18657511]
- Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc*. 2009; 4:1471–1501. [PubMed: 19798082]
- Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A*. 2011; 108:2623–2628. [PubMed: 21262818]
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*. 2011; 29:143–148.
- 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. [PubMed: 21421566]

- Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science*. 2009; 326:1501. [PubMed: 19933106]
- Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research*. 2011; 39:9283–9293. [PubMed: 21813459]
- Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*. 2012 Describes archive of plasmids encoding pre-assembled TALE repeat arrays that is required to practice REAL-Fast and demonstrates that TALENs constructed on the same architecture as those made by REAL and REAL-Fast function robustly and have high rate of success in human cells.
- Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, Yeh JR. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature biotechnology*. 2011; 29:697–698. Describes the plasmids required to practice REAL and the efficacy of TALENs produced by this architecture in zebrafish.
- Scholze H, Boch J. TAL effectors are remote controls for gene activation. *Curr Opin Microbiol*. 2011
- Sun N, Liang J, Abil Z, Zhao H. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst*. 2012; 8:1255–1263. [PubMed: 22301904]
- Tesson L, Usal C, Menoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, et al. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol*. 2011; 29:695–696. [PubMed: 21822240]
- Weber E, Gruetzner R, Werner S, Engler C, Marillonnet S. Assembly of designer TAL effectors by Golden Gate cloning. *PLoS ONE*. 2011; 6:e19722. [PubMed: 21625552]
- Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, et al. Targeted genome editing across species using ZFNs and TALENs. *Science*. 2011; 333:307. [PubMed: 21700836]
- Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology*. 2011; 29:149–153.

**A.**



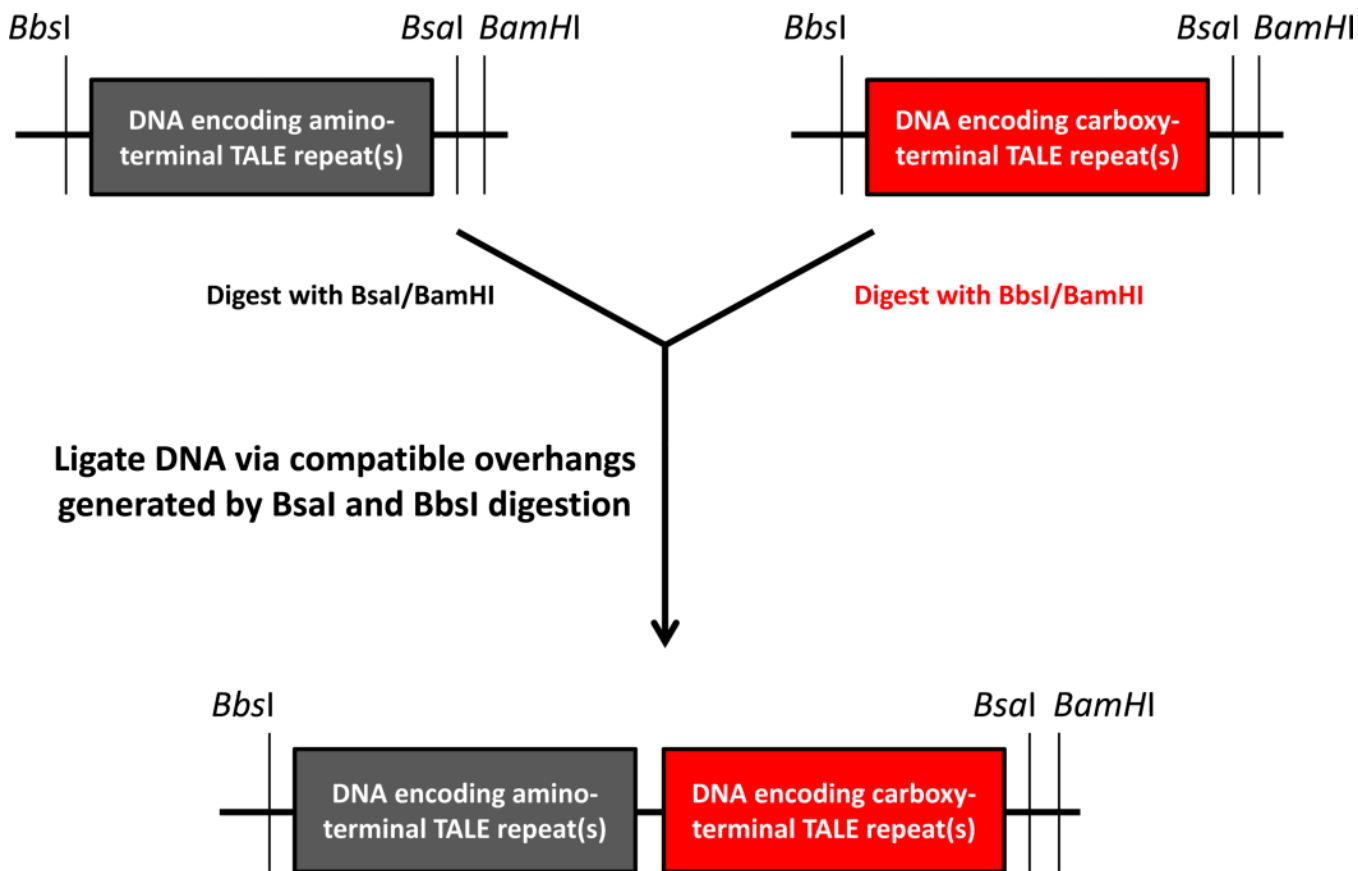
**B.**



**Figure 1. Engineered TALE Nucleases**

(A) Schematic illustrating engineered TALE nucleases. Colored rectangles represent individual TALE repeat domains with the identities of the repeat variable di-residues (RVDs) shown. FokI nuclease domains are represented as purple octagons. TALENs bind as dimers with each monomer binding a “half-site” (grey rectangles) and cleavage occurring in the “spacer” sequence (white rectangle) between the two half-sites.

(B) Assembly of individual TALE repeat domains into engineered TALE repeat arrays. TALE repeat domains (colored rectangles) with different RVDs bind to different single bases. Assembly of individual repeats generates arrays capable of binding to extended DNA sequences.



**Figure 2. Overview of ligation strategy used to join together DNA encoding TALE repeats**  
 A fragment of DNA encoding carboxy-terminal TALE repeat(s) created by digestion with *BbsI* and *BamHI* restriction enzymes is ligated to a vector backbone of DNA encoding amino-terminal TALE repeat(s) created by digestion with *BsaI* and *BamHI*. Ligation occurs via compatible DNA overhangs created by the TypeIIS enzymes *BbsI* and *BamHI*. The resulting plasmid can be used in subsequent iterative ligations of the same type shown here.

ZINC FINGER CONSORTIUM

ZiFiT Targeter Version 4.1

Introduction ZiFiT Instructions Examples FAQ References Funding Links

Please indicate with brackets the nucleotide at which you would like the double-stranded DNA break (DSB) to occur  
For example: AGCTACGATCGACTACGATAGTCGGCTACGGACTGAC[T]GACTGACTAGCTACGATCGACT

Sequence:

>Sequence  
AGCTACGATCGACTACGATAGTCGGCTACGGACTGAC[T]GACTGACTAGCTACGATCGACT

Only unique sites can be displayed when there are multiple queries

Mask redundant sites  Relax Constraints (see FAQ)

Submit

SEQ 1

Query Sequence ( ) : \*

cAGCTACGATCGACTACGATAGTCGGCTACGGACTGACTGACTAGCTACGATCGACT

Tier One:  
[TCGACTACGATAGTCGGCTACGGACTGACTGACTAGCTACGATCGACT](#)

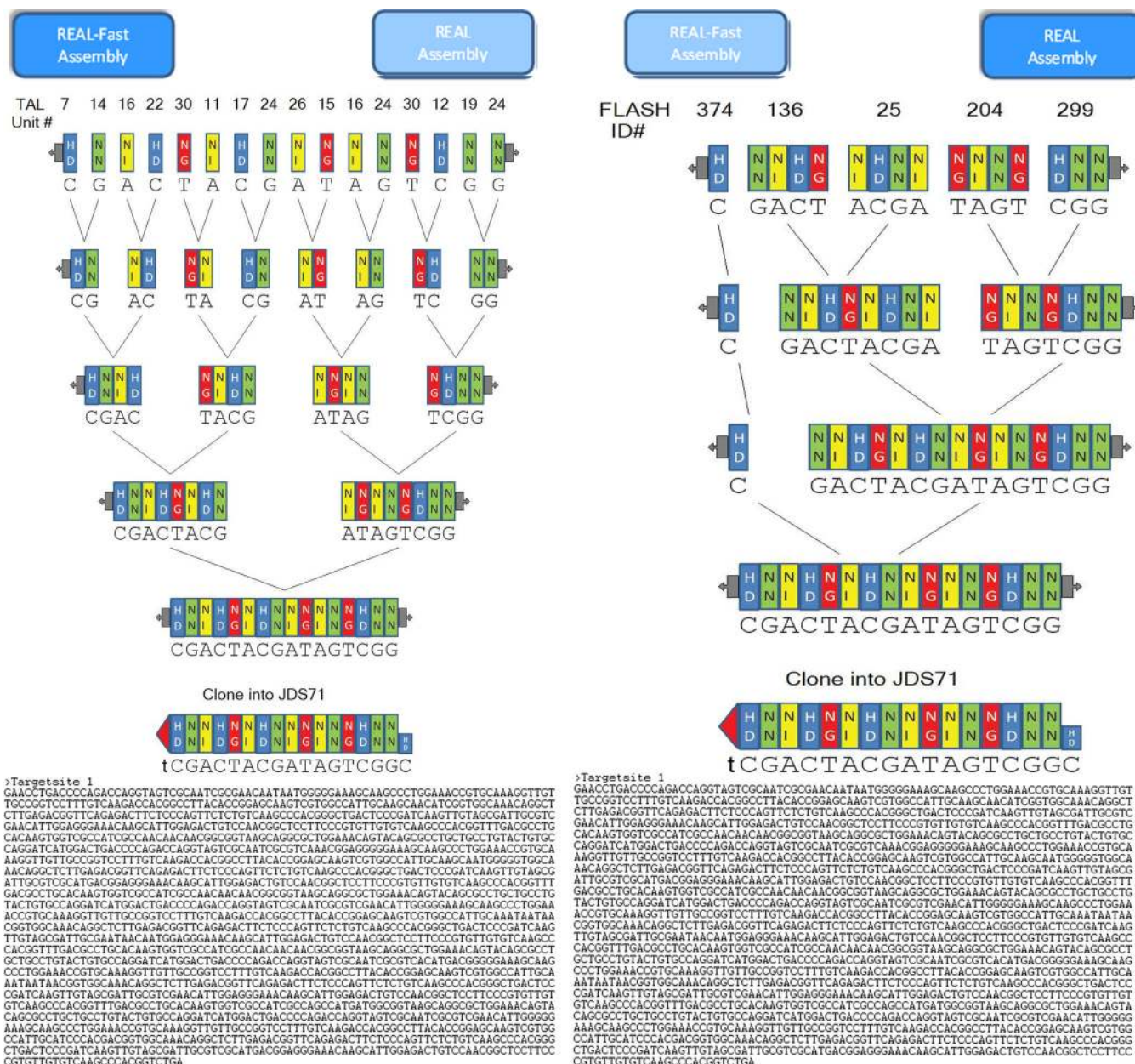
Tier Two:  
[TACGATCGACTACGATAGTCGGCTACGGACTGACTGACTAGCTACGA](#)

(Click on the target site for a graphical guide summarizing construction steps)

### Figure 3. ZiFiT Targeter output

A screenshot of a typical output from ZiFiT Targeter is shown. In the example shown, a query sequence has been submitted and an output with five potential target sites has been returned below. Note that for reference the original sequence is shown in the output as a “Query Sequence” with the original bracketed nucleotide marked with an asterisk. In the example shown, two of the sites are “First Tier” sites and three are “Second Tier” sites (see text for additional details). For each target site, the half-sites are shown in blue text while the spacer sequence is shown in red text. The target half-sites are hyperlinks that can be clicked to open a new page with a customized graphical guide for assembling the TALEN that binds to that half-site. As described in the text, users can also choose to relax search criteria further to allow identification of additional potential TALEN target sites by checking the box for “Relax Constraints” and/or unchecking the box for “Mask Redundant Sites” and then clicking the Search button again.





**Figure 4. Example of a customized graphical guides generated by ZiFiT Targeter for construction of an engineered TALEN**  
 This particular guide illustrates the construction of a TALE array consisting of 15.5 repeats. TALE repeats are represented as colored rectangles with the identities of their RVDs shown in two letter code within the rectangle. Bases bound by each TALE repeat are shown beneath the rectangle. Ligations of units to be performed are indicated by black lines. The numbers of plasmids used to perform the initial ligations are shown above the first row. The final assembled array is cloned into the pJDS-series TALEN expression vector indicated at the bottom. Cloning into this vector adds the required carboxy-terminal 0.5 TALE repeat domain and fuses the assembled TALE repeat array to the wild-type FokI nuclease domain, thereby generating a TALEN expression plasmid. A) Graphic for TALEN assembled using REAL B) Graphic for TALEN assembled using REAL-Fast