Fully Automated One-Step Synthesis of Single-Transcript TALEN Pairs Using a Biological Foundry

Ran Chao, Jing Liang, Ipek Tasan, Tong Si, Linyang Ju, and Huimin Zhao

1Carl R. Woese Institute for Genomic Biology, 2Department of Chemical and Biomolecular Engineering, 3Department of Biochemistry, 4Department of Microbiology, 5Department of Chemistry, and 6Department of Bioengineering, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States

ABSTRACT: Transcription activator-like effector nuclease (TALEN) is a programmable genome editing tool with wide applications. Since TALENs perform cleavage of DNA as heterodimers, a pair of TALENs must be synthesized for each target genome locus. Conventionally, TALEN pairs are either expressed on separate vectors or synthesized separately and then subcloned to the same vector. Neither approach allows high-throughput construction of TALEN libraries for large-scale applications. Here we present a single-step assembly scheme to synthesize and express a pair of TALENs in a single-transcript format with the help of a P2A self-cleavage sequence. Furthermore, we developed a fully automated platform to custom manufacture TALENs in a versatile biological foundry. 400 pairs of TALENs can be synthesized with over 96.2% success rate at a material cost of $2.1/pair. This platform opens the door to TALEN-based genome-wide studies.

KEYWORDS: TALEN, DNA assembly, genome engineering, genetic screening, biofoundry, automation

T

ranscription activator-like effector nuclease (TALEN) is a highly efficient and programmable genome editing tool, which has been applied in a wide range of organisms. A TALEN consists of a FokI DNA cleavage domain and a DNA binding domain (DBD) that has tandem repeats of a 33–35 amino acids (aa) motif. The 12th and 13th aa residue within each repeat is known as repeat-variable diresidue (RVD), and it determines the DNA binding specificity of the repeat. By assembling repeats with specific RVDs in order, a TAL effector DBD can bind to a specific DNA sequence. Because the FokI cleavage domain functions as a dimer, TALENs are typically used in tail-to-tail heterodimeric pairs to create double stranded breaks for genome editing. Such heterodimeric design generates high editing efficiency and improves specificity, but it also presents challenges in TALEN synthesis as well as usage.

A number of methods have been developed to synthesize TALEN expression DNA vectors. Taking advantage of an optimized set of 4-bp junctions as well as a preassembled direpeat part library, we developed a one-step assembly scheme based on the Golden Gate method. Custom TALEN vectors can be constructed in 24 h at 96% success rate and a material cost of $5. These methods, however, can only assemble vectors harboring a single TALE-FokI monomer. Since TALEN requires a heterodimer to make a cut, two monomers must be introduced into the host cells either on two separate vectors or a single subcloned vector with both monomers. Either option has significant drawbacks. For example, both of them will require twice as many vectors synthesized as the number of target sequences. When the monomers are on separate vectors, the number of cells transfected or transformed with both monomers can be reduced. More importantly, the dual vector scheme makes it very difficult to perform high-throughput genetic screening. Thanks to fluorescence-activated cell sorting (FACS) and next-generation sequencing, a large number of cells with different genotypes can be screened for phenotypes of interest and sequenced. As a precision genome editing tool, TALEN can potentially be used to generate a genomic knockout library. However, because the two monomers of each TALEN pair need to be introduced to the same cell, library transfection or transformation is not possible using a dual vector system. However, current methods to construct a single-vector TALEN require a lengthy and complicated subcloning procedure, which makes the synthesis process difficult to scale up. A high-throughput synthesis method for single-vector TALENs will open up new possibilities.

Building on our previously published “FairyTALE” protocol, we sought to assemble a pair of TALEN monomers onto a single vector in a one-step reaction. In a previous work, 2A self-cleavage peptide was used to cotranscribe a pair of TALENs as one mRNA molecule but translated as separate functional proteins. We operationalized this coexpression strategy in a 15-insert one-pot assembly scheme, and assembled single-plasmid TALENs in one step at more than 87.7% fidelity.

Received: October 12, 2016
Published: January 19, 2017

DOI: 10.1021/acssynbio.6b00293
ACS Synth. Biol. 2017, 6, 678−685
TALENs synthesized using this single-transcript design had comparable cleavage activity in mammalian cells as those synthesized using a two-plasmid design. We implemented the synthesis on iBioFAB (Illinois Biofoundry for Advanced Biomanufacturing), an integrated and versatile robotic system, to fully automate the synthesis process. 400 pairs of TALENs can be generated on a daily basis at a material cost of $2.1 per pair with minimal human intervention. We envision that genome-wide studies using TALENs can be scaled up to screen hundreds of loci in parallel with such a simplified design and automated synthesis.

■ RESULTS AND DISCUSSION

Design of a single-transcript TALEN synthesis scheme. The TALEN architecture used in this work is based upon the AvrXa10 TALE from Xanthomonas oryzae pv oryzae, as previously reported. In brief, it utilizes a +207 aa N-terminus extension and a +63 aa C-terminus extension, which negates the 5′-T requirement and allows greater flexibility in target sequence design. Attached at the C-terminus is an engineered FokI cleavage domain that showed greater cleavage efficiency in yeast as well as human cells. The central repeat domains of the two TALENs are constructed from a library of direpeat substrates; that is, each substrate contains two TALE repeats that recognize two consecutive DNA bases. For this work, we used a library of 441 direpeat substrates, adapted from “FairyTALE”, divided into 17 groups according to their position in the assembly (Supplementary Table 1). In addition to the 42 substrates to cover all possible DNA dibases at each assembly position, we included the option to use either NH or NN to code for guanine. To separate the two TALENs on the single plasmid, we employed a polycistronic format utilizing a P2A self-cleaving peptide sequence. Both TALENs are coded as a single transcript, but during translation, the P2A peptide will self-cleave the growing polypeptide to give two independent TALENs (Figure 1a).

Using the set of optimized 4-bp junctions in the “fairyTALE” construction scheme, 2 sets of 7 direpeat substrates, with a P2A linker substrate in between, were ligated onto a TALEN receiver vector in a single step via Golden Gate assembly. The N-terminus extension of the first TALEN and the C-terminus extension of the second TALEN were carried by the vector, whereas the C-terminus extension of the first TALEN and the N-terminus extension of the second TALEN were carried by the linker substrate. Since the linker substrate and the receiver carried the last repeat of the two TALENs, 4 TALEN receivers and 4 linker substrates were created (Figure 1b). This construction scheme assembled 15 DNA fragments onto a 5 kb mammalian expression vector to create a single-plasmid TALEN pair that recognized a 30 bp DNA sequence.
One-pot assembly of TALENs. To fulfill the requirements of TALEN library creation, we optimized the reaction condition to maximize the assembly fidelity. For library creation applications, picking individual clones for verification would be an obvious throughput bottleneck, and we would therefore need to achieve high assembly fidelity to allow us to skip clonal isolation without drastically affecting the quality of the library. We picked 28 colonies from a single-transcript TALEN assembly, and assessed them by restriction digest followed by gel electrophoresis. As shown in Figure 1c, all 28 clones gave the correct digestion pattern. We then sequenced 4 of the clones, and they all appeared to be correct. This (28/28) corresponds to a fidelity of at least 87.7% based on binomial probability with 95% confidence.

Single-transcript TALEN functionality in HEK293T and hESC cells. To ensure that P2A cleaves the protein effectively, we performed a Western blot analysis from the cell lysates of HEK293T that had been transfected with single-plasmid TALENs. As shown in Figure 2a, only TALEN monomer was detected and no dimer could be observed, suggesting that the P2A sequence cleaved the protein effectively in HEK293T cells.

After confirming P2A functionality, we went on to compare the DNA cleavage efficiency of single-transcript TALENs against previously reported traditional two-plasmid TALENs. Two sites, ABL1 and BRCA2, were chosen for this comparison, and the experiments were performed in HEK293T cells. Cleavage efficiency was measured using the T7E1 nuclease assay, which detects indels introduced via NHEJ after TALEN induced double stranded breaks. As shown in Figure 2b, the cleavage efficiency of the two single-transcript TALENs was comparable to that of traditional TALENs. The 1P-TALEN used in this experiment used NH to recognize guanine, whereas the traditional TALENs used NN to recognize guanine. According to our observation and in agreement with that reported by others, when used in large number, NH RVD is detrimental to TALE binding. We therefore recommend using NN or a mix of NN and NH RVD when there are more than 4 guanine bases in the recognition site (Supplementary Figure 2).

Full automation of single-transcript TALEN assembly on a biological foundry. Automation has been used to accelerate biological engineering by either reducing human interventions in individual steps, or completely eliminating human intervention using integrated systems. The latter approach has demonstrated the great power of full automation by creating a large number of genetic variants in a short time period. To enable large-scale applications of TALENs, such as genetic screening, we sought to fully automate the synthesis process of TALENs. However, existing integrated platforms are extensively customized for specific tasks and difficult to reconfigure. It would not be efficient and economical to build a deeply customized system dedicated to TALEN synthesis.

Figure 2. Functional test of single-transcript TALENs. (a) Single-transcript expression of a TALEN pair. Two distinctive TALEN pairs were expressed in HEK293T cells with the single-transcript design. TALEN monomers showed visible bands on Western blot while no band for the size of the uncleaved doublet was detected. (b) Genome editing in HEK293T cells. Single-transcript TALENs (STTLN) were compared against the traditional dual plasmid TALENs (TDTLN) by targeting BRCA2 as well as ABL1 sites in HEK293T cells. T7E1 assay was performed to detect the indel introduced by TALEN cleavage and NHEJ. The STTLN transfected samples showed comparable cleavage efficiency to TDTLN transfected samples. CTRL: sample with no TALEN transfection served as negative control. (c) Disruption of an Oct4 enhancer in H1 hESC. Flow cytometry was used to quantify the GFP expression level in H1-Oct4-GFP cells. The gated population had lower than normal GFP expression. Left: control population without enhancer disruption, middle: enhancer disrupted by traditional 2-plasmid TALEN, right: enhancer disrupted by single-transcript TALEN.
Instead, we applied a generalized Golden Gate assembly workflow implemented on iBioFAB. The iBioFAB consisted of component instruments, a central robotic platform, and a modular computational framework (Figure 3). Twenty devices, each in charge of a unit operation, such as pipetting and incubation, were linked by two robotic arms into various process modules, such as DNA assembly and transformation, and then further organized into workflows such as pathway construction and genome engineering (Figure 3a,b). An overall scheduler was developed to orchestrate the unit operations and allow hierarchical programming of the workflows (Figure 3c). The iBioFAB was configured to perform a generalized automatic DNA assembly workflow where various kinds of DNA constructs can be manufactured on-demand with the Golden Gate method.13 A sequence of unit operations was designed to implement this workflow (Figure 4a,b). To streamline the process, we developed Script Generator, a design tool that automatically converts DNA assembly designs to experimental routines of mix-and-matching arbitrary DNA parts. Script Generator then generates robotic commands for iBioFAB to conduct the complex pipetting work. The pipetting routes were also optimized to minimize tip and time consumption. The aspiration steps are combined as much as possible for the same substrate and dispensed into the corresponding destination. Tips are loaded on demand from the storage carousel to the liquid handling station.

In this work, we adapted this DNA assembly workflow for synthesizing single-transcript TALENs. An extension that automated the DNA assembly design specifically for TALENs was added to Script Generator. Using such a pipeline, the operator only needs to input the target DNA sequence to Script Generator, and iBioFAB would perform the rest of TALEN synthesis with minimal human intervention. It only requires the operator to load reagents and consumables on a daily basis. Any arbitrary number between 1 to 192 TALEN pairs can be synthesized in each batch.

High-throughput synthesis of 192 single-transcript TALENs. To test the high-throughput synthesis pipeline, we fed 192 different human genomic target loci to Script Generator. iBioFAB performed 3648 pipetting steps from 444 different DNA parts, and reagents within 17 h at a material cost of $2.1 per TALEN pair (Supplementary Table 2). By staggering batches, over 400 TALENs can be generated in a single day (Supplementary Figure 3).

To evaluate the success rate of the synthesis, 94 randomly selected constructs were verified by polyclonal restriction digestion. All samples showed the correct digestion pattern (Figure 4c), which corresponds to a success rate of at least 96.2% with 95% confidence based on binomial probability. For activity verification, we randomly selected 22 TALENs for T7E1 assay in HEK293T cells.14 Fifteen of 22 samples showed cleavage activity. Since cleavage activity was known to be sequence dependent, the lack of activity for some sites was not unexpected.15 To eliminate the possibility of misassembly, we sequenced all the constructs that did not show cleavage activity. All sequencing reads were aligned to the intended TALEN designs, indicating that the TALENs were correctly assembled (Supplementary Table 3).

Besides TALEN, CRISPR-Cas9 is another popular technology used in genome editing applications.16 As opposed to using a specific protein to recognize DNA sequences, CRISPR utilizes RNA to perform the recognition through base pairing. Using a...
nucleic acid for targeting has many advantages, but most importantly, through the use of microarray DNA synthesis, a large nucleic acid library is readily accessible. As such, even though TALEN had a two-year head start over CRISPR, multiple targeting and genetic screening were both first achieved using CRISPR.\(^6\),\(^{17}\) However, due to its relatively short recognition sequence, 20 bp, the off-target effect is a significant problem in CRISPR.\(^{18}\) In a genetic screening that targets structural genes, the off-target effect can be compensated by targeting multiple sites within the same gene, so that a high-confidence hit can be identified by looking for the enrichment of a set of sites instead of any single site. However, in the case where the functional DNA element is very small, e.g., a transcriptional enhancer, or a miRNA gene, there is simply not enough length to fit in multiple targeting sites. Furthermore, in the case where the functional DNA element is very small, e.g., a transcriptional enhancer, or a miRNA gene, there is simply not enough length to fit in multiple targeting sites. Given the limited range for target selection, CRISPR may not be able to find a site that is sufficiently unique in the genome. Furthermore, given the small number of selectable sites for such screens, the level of confidence for any resultant hits will be low. A TALEN library, with a different off-target profile, can be used in conjunction with a CRISPR library to improve the confidence of any potential hits.

In conclusion, we have developed a scheme to synthesize TALEN pairs on a single vector in a one-pot reaction, which has substantially simplified the synthesis of TALENs while achieving an outstanding success rate. An automated process was developed accordingly, and the resulting pipeline makes it possible to create large TALEN libraries at a reasonable cost and in a reasonable time frame.

**METHODS**

iBioFAB. iBioFAB consists of a F5 robotic arm on a 5-m track (Fanuc, Oshino-mura, Japan), an Evo200 liquid handling robot (Tecan, Männedorf, Switzerland), two shaking temperature controlled blocks (Thermo Scientific, Waltham, MA), a M1000 microplate reader (Tecan, Männedorf, Switzerland), a Cytomat 6000 incubator (Thermo Scientific, Waltham, MA), two Cytomat 2C shaking incubators (Thermo Scientific, Waltham, MA), three Multidrop Combi reagent dispensers (Thermo Scientific, Waltham, MA), four Trobot thermocyclers (Biometa, Göttingen, Germany), Vspin plate centrifuge (Agilent, Santa Clara, CA), a storage carousel (Thermo Scientific, Waltham, MA), a light box, and a camera for colony picking (Scirobotics, Kfar Saba, Israel).

**Figure 4.** Fully automated synthesis of TALEN libraries. (a) General workflow for the DNA assembly pipeline based on the Golden Gate method. Script Generator converted project design ideas, such as permutations of DNA parts, to assembly designs with appropriate extensions and further robotic commands for pipetting the stock plasmids to DNA mixes. In Golden Gate reactions, Type IIs restriction enzymes such as BsaI generated a set of standard precharacterized 4-bp single strained ends as linkers. The corresponding linkers annealed and were ligated by T4 ligase. (b) Process flow diagram for the build step. Unit operations employed were marked in blue. (c) Verification of single-transcript TALENs synthesized in high throughput. 94 samples were randomly selected from the 192 TALEN pairs synthesized in the full batch test. Each plasmid sample encoding a pair of TALENs was extracted from a polyclonal E. coli cell culture and restriction digested. The fragment sizes were analyzed by capillary electrophoresis. The digestion pattern was simulated. (d) Summary of efficiency, throughput, and cost.
between unit operations. Freedom Evoware (Tecan, Männedorf, Switzerland) was used to control the liquid handling robot and program pipetting modules. Pipetting modules specifically defined the general procedure of pipetting on the liquid handling robot, such as labware fetching from the central robotic arm, DNA part dispensing, reagent dispensing, and temperature controls. iScheduler and ScriptGenerator are programmed in Visual Basic. iScheduler executes process modules by sending commands in Extensible Markup Language to Momentum. The ScriptGenerator converted user defined DNA assembly as permutations of parts to source and destination locations based on preloaded parts storage plate layouts. The corresponding pipetting routines were optimized by queueing the destination locations from the same source. Pipetting worklists were compiled accordingly and sent to Freedom Evoware to control aspiration, dispense, as well as tip change actions. Defined amount of each DNA part was aspirated and multidispensed without contacting the liquid in the destination. Tips were reused as much as possible and changed when all destinations for the same source were dispensed. Constraints such as tip volume and maximum number of aspirations with each tip were also imposed in the algorithm.

**Plasmids.** Based on the RVDs parts used in the previous work, a new library of TALEN stock plasmids were developed for the single plasmid design. The group for position 6 was replaced with LR_N-term_FokI_P2A+C-term constructs (Supplementary Figure 1a). Dual and single RVD parts with NN were supplemented into the stock library. The RVD and P2A fragments were inserted to a receiver plasmid (Supplementary Figure 1b) with human CMV promoter as well as last P2A fragments were inserted to a receiver plasmid (Supplementary Figure 1b) with human CMV promoter as well as last

**Golden gate assembly and verification.** Golden Gate DNA assembly was performed with the methods described in the previous work. Competent *E. coli* HST08 strain (Clontech, Mountain View, CA) was prepared with Mix & Go *E. coli* Transformation Buffer Set (Zymo Research, Irvine, CA). 2.5 μL of Golden Gate reaction products were first mixed with *E. coli* competent cells on a Peltier block held at 0 °C and incubated for 30 min. The cell plate was then transferred to a second Peltier block held at 42 °C by the plate manipulation arm. After 1 min heat shock, the cell plate was transferred back to the 0 °C block and chilled for 2 min. The transformants were recovered in LB broth (Becton, Dickinson and Company, Franklin Lakes, NJ) for 1 h. The recovered cell suspensions were either plated on LB agar media with 100 μg/mL of ampicillin or used to inoculate polyclonally LB liquid media supplemented with 200 μg/mL of carbenicillin. Plasmids were extracted from the polyclonal cultures with MagJET Plasmid DNA Kit (Thermo Scientific, Waltham, MA) and restriction digested by EcoRI-HF (New England Biolabs, Ipswich, MA). The digestion products were analyzed by 1% agarose gel in low throughput or Fragment Analyzer (Advanced Analytical Technologies, Ankeny, IA) in high throughput. Selected plasmids were also verified by Sanger sequencing reactions (ACGT, Wheeling, IL) with 4 primers. The binomial probability confidence interval for assembly success rate was calculated with Clopper-Pearson method.

**Mammalian gene knockout and verification.** Human embryonic kidney (HEK) cell line HEK293T was transfected with randomly selected TALEN plasmids. HEK293T cells were used as they are easy to cultivate and transfect. Although no cell authentication or mycoplasma contamination tests were performed, we reason that the results of T7E1 assay is relatively insensitive to the cell line background. Cells were maintained in Dulbecco’s modified Eagle’s Medium (DMEM) (Corning Life Sciences, Tewksbury, MA) supplemented with 10% heat inactivated fetal bovine serum (Life Technologies, Carlsbad, CA) at 37 °C and 5% CO2 incubation. One day prior to transfection, 293T cells were seeded into 12-well BioCoat Collagen-I coated plates (Corning Life Sciences, Tewksbury, MA) at a confluency of ~50%. Transfections were performed with FuGENE HD Transfection Reagent (Promega, Madison, WI) according to the manufacturer’s protocols. Briefly, for each well of the 12-well plate, 1 μg of clonally purified TALEN plasmid was first diluted in Opti-MEM (Life Technologies, Carlsbad, CA) to a total volume of 100 μL. After addition of 3 μL FuGENE HD reagent and incubation at room temperature for 5 min, the mix was added onto the cells. Cells were harvested at 60 h post-transfection. The genomic DNA was extracted with QuickExtract DNA Extraction Solution (Epicenter, Madison, WI).

The cleavage efficiency was evaluated by T7E1 assay. DNA amplicons were designed to have a length of 400–1000bp flanking the nominal cleavage site by a custom developed Visual Basic script. It searches the genome sequence within a given range for a pair of primer binding sites to avoid off-targets, long stretches of GC, AT, or any single type of nucleotide. End nucleotides, GC contents, and melting temperatures were optimized. The relevant genome sequences were downloaded by querying UCSC DAS server (https://genome.ucsc.edu/cgi-bin/das/) while off-target check was performed by querying GGGenome server (https://gggenome.dbcls.jp/). The PCR amplification was conducted with Q5 polymerase (New England Biolabs, Ipswich, MA) and annealing temperature touchdown (65–55 °C for 10 cycles, 55 °C for 20 cycles). In the cleavage assay, 200 ng of purified amplicon in 10 μL NEB Buffer 2 was first denatured and refashioned (95 °C, 5 min; 95–85 °C at ~2 °C/s; 85–25 °C at ~0.1 °C/s; hold at 4 °C). 10U of T7 Endonuclease I (New England Biolabs, Ipswich, MA) was added and incubated at 37 °C for 15 min. The reaction was stopped by adding 1 μL of 0.5 M EDTA. The digestion products were analyzed by Fragment Analyzer (Advanced Analytical Technologies, Ankeny, IA).
ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.6b00293.

Additional figures and tables as described in the text (PDF)

AUTHOR INFORMATION

Corresponding Author
*Phone: (217) 333-2631. Fax: (217) 333-5052. E-mail: zhao5@illinois.edu.

ORCID
Huimin Zhao: 0000-0002-9069-6739

Author Contributions
R.C. and J.L. contributed equally. R.C., J.L., and H.Z. designed and conceived the study. R.C. developed the iBioFAB automation system and automated the DNA assembly workflow. J.L. developed the P2A-based TALEN synthesis scheme, performed initial tests, and selected genome editing targets. I.T. performed TALEN functional tests in human cells. L.J. helped with preparation of TALEN plasmids and targets. I.T. performed TALEN functional tests in human cells. R.C., J.L., and H.Z. wrote the manuscript with important input from T.S. and I.T.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS
This work was supported by the Roy J. Carver Charitable Trust, Carl R. Woese Institute for Genomic Biology at the University of Illinois at Urbana-Champaign (UIUC), Defense Advanced Research Projects Agency, and National Institutes of Health (1U54DK107965). R.C. acknowledges fellowship support from 3M Corporation. T.S. acknowledges postdoctoral fellowship support from Carl R. Woese Institute for Genomic Biology (UIUC). The authors acknowledge Vasundhara Vigh for her support from Carl R. Woese Institute for Genomic Biology (1U54DK107965). R.C. acknowledges fellowship support from Carl R. Woese Institute for Genomic Biology (1U54DK107965). R.C. and J.L. contributed equally. R.C., J.L., and H.Z. designed and conceived the study. R.C. developed the iBioFAB automation system and automated the DNA assembly workflow. J.L. developed the P2A-based TALEN synthesis scheme, performed initial tests, and selected genome editing targets. I.T. performed TALEN functional tests in human cells. L.J. helped with preparation of TALEN plasmids and verification of genome editing efficiency. H.Z. supervised the research. R.C., J.L., and H.Z. wrote the manuscript with important input from T.S. and I.T.

ABBREVIATIONS
TALEN, transcription activator-like effector nucleases; CRISPR, clustered regularity interspaced short palindromic repeat; iBioFAB, Illinois Biological Foundry for Advanced Bioengineering

REFERENCES


