**SunnyTALEN: A Second-Generation TALEN System for Human Genome Editing**

Ning Sun,1 Zehua Bao,1 Xiong Xiong,2 Huimin Zhao1,2,3

1Department of Biochemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801; telephone: +217-3332631; fax: +217-3335052; e-mail: zhao5@illinois.edu
2Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801
3Department of Bioengineering, Department of Chemistry, Center for Biophysics and Computational Biology and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

**ABSTRACT:** Transcription activator-like effector nucleases (TALENs) have rapidly emerged as a powerful genome editing tool. The site-specific DNA double-strand breaks generated by TALENs in the human chromosome can induce homologous recombination or non-homologous end joining, resulting in desired genetic modifications. In this study, we report the development of a TALEN variant, SunnyTALEN, with >2.5-fold improved genome editing efficacy in human cells. The corresponding scaffold increases the rate of genetic modification at all the 13 tested loci of human genome and is compatible with heterodimer TALEN architectures. This enhanced and high-efficiency TALEN variant represents a novel second-generation TALEN system and has great potential for biological and therapeutic applications.


**KEYWORDS:** directed evolution; double-strand break; genome editing; gene therapy; synthetic biology

**Introduction**

Genetic modifications of human genomes have a wide range of applications such as investigating human biology, studying disease mechanisms and treating diseases via gene therapies (Cherry and Daley, 2013; Dow and Lowe, 2012). Although virus-based transgenesis can modify human genomes efficiently by introduction of exogenous genes, random integration of viral transgenes poses the risk of insertional mutagenesis and carcinogenesis (Check, 2002; Marshall, 1999). In contrast, targeted genome editing or engineering enables researchers to tailor the human genome in a precise manner. To achieve efficient genome editing in human cells, DNA double-strand breaks (DSBs) have to be introduced at the specific sites of the chromosome by custom-designed DNA endonucleases. Subsequent repair of the DSBs by non-homologous end joining (NHEJ) or homologous recombination can generate desired genomic modifications such as gene disruptions, gene deletions, gene insertions, gene replacements, and chromosome rearrangements (Perez-Pinera et al., 2012; Sun et al., 2012a).

Recently, transcription activator-like effector (TALE) nucleases (TALENs) have emerged as an efficient and versatile tool for genome editing by introduction of chromosomal DNA DSBs (Christian et al., 2010; Joung and Sander, 2013; Miller et al., 2011; Mussolino and Cathomen, 2012; Sun and Zhao, 2013b). TALENs are artificial DNA endonucleases composed of a TALE central repeat domain as the DNA recognition module and a non-specific FokI nuclease domain as the DNA cleavage module. The TALE central repeat domain comprises a series of tandem repeat units, each of which typically contains 34 residues. Each repeat is used to recognize a single nucleotide and the DNA recognition specificity is determined by the highly variable residues at positions 12 and 13 (e.g., NI recognizes adenine, HD recognizes cytosine, NG recognizes thymine, NH recognizes guanine, and NN recognizes guanine and adenine) (Boch et al., 2009; Cong et al., 2012; Moscou and Bogdanove, 2009; Streubel et al., 2012). The simple DNA recognition code and the modular nature of the TALE central repeat units enable researchers to tailor DNA recognition specificity with ease and target essentially any
desired DNA sequence. Therefore, TALENs have been widely applied for genome editing in various species (Sun and Zhao, 2013b). Due to their long recognition sequences (30–50 bp) and high DNA recognition specificity, TALENs exhibit minimal off-target effects in the context of complex human genomes (Ding et al., 2013; Mussolino et al., 2011; Sun et al., 2012b), making them an ideal platform for human genome editing.

First-generation TALEN scaffolds have been described with different N-terminal segments (NTSs) and C-terminal segments (CTSs) flanking the TALE central repeat domains (Kim et al., 2013a; Miller et al., 2011; Mussolino et al., 2011; Sun et al., 2012b). Some of them have been applied for generating human stem cell-based disease models (Ding et al., 2013) and treating human diseases (Choi et al., 2013; Osborn et al., 2013; Sun and Zhao, 2013a), but their efficacy of modifying human genomes is limited in targeting certain loci (Kim et al., 2013a; Reyon et al., 2012). Although a second-generation TALEN platform called GoldyTALEN has been reported with improved genome editing efficacy in zebrafish (Bedell et al., 2012) and livestock (Carlson et al., 2012), its efficacy in modifying human genomes has not been demonstrated. Here, we report the development and application of a high-throughput screening system to improve TALEN activity through a directed evolution strategy. We successfully isolated a TALEN variant, SunnyTALEN, with significantly improved efficacy in modifying human genomes. We demonstrate that the corresponding TALEN scaffold is more active than the GoldyTALEN system in human cells and compatible with previously published obligate heterodimeric FokI nuclease domains. This novel second-generation TALEN system provides a general solution for efficient modifications of human genomes with low toxicities and has great potential in both basic and applied biological sciences.

Materials and Methods

Materials

Q5 DNA polymerase, T4 DNA ligase, Antarctic phosphatase, and restriction endonucleases were purchased from New England Biolabs (Beverly, MA). QIAprep Spin Plasmid Miniprep Kit, QIAquick Gel Extraction Kit, and QIAquick PCR Purification Kit were obtained from Qiagen (Valencia, CA). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). All the other reagents unless specified were obtained from Sigma-Aldrich (St. Louis, MO).

Yeast Reporter Strain

To construct a single-strand annealing enhanced green fluorescent protein (eGFP) reporter, two separated eGFP fragments sharing a 100-bp homologous region were PCR amplified separately, digested with AvrII and XbaI, respectively, and ligated overnight at 16°C. The ligation product was subsequently digested with a combination of AvrII and XbaI and the 870-bp non-cut fragment was gel-purified. The eGFP reporter gene was then cloned into the pRS414 plasmid (New England Biolabs) between the ADH1 promoter and the ADH1 terminator. The centromere sequence of the plasmid was then removed by the PmlI digestion and blunt ligation. This plasmid was then digested with HindIII at the TRP1 gene. The linearized plasmid was transformed into *Saccharomyces cerevisiae* HZ848 (MATa, ade2-1, ade3Δ22, Δura3, his3-11,15, trp1-1, leu2-3,112, and can1-100) using the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl, 2007) and stably integrated into the chromosome at the *TRP1* site. The yeast reporter strain was obtained by selection of the transformed yeast cells on plates containing synthetic complete medium lacking tryptophan with 2% glucose at 30°C. The integration was confirmed by PCR amplification and DNA sequencing of the targeted genomic region.

Library Creation

The pRS415 yeast expression vector encoding the AvrXa10 TALEN that has been described previously (Sun et al., 2012b) was used as the template for mutagenic PCR. The sequence encoding the CTS and FokI nuclease domain was mutagenized using the GeneMorph II Random Mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer’s instruction. The following primers were used for the amplification: EP-FokI-Far-for 5'-tgg cct act gca ggta ctg-3' and EP-FokI-Rev-ref 5'-cgt gaa act tcg aac act gtc-3'. Two libraries with a low mutation rate (0–4.5 mutations/kbp) and a medium mutation rate (4.5–9 mutations/kbp) were amplified and gel-purified, respectively. The TALEN expression plasmid was digested with *AarII* and *SalI* and gel purified to remove the sequences encoding the wild-type CTS and FokI domain. The two gene libraries were mixed and co-transformed with the linearized TALEN expression plasmids at an approximately 13:1 insert:vector molar ratio into the yeast reporter strain using the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl, 2007). Through yeast homologous recombination, the pRS415 plasmid library was created containing ~2 × 10^5 transformants.

High-Throughput Screening

After transformation, the yeast reporter strain carrying the library of TALEN variants was recovered at 30°C with shaking for 1 h in YPA medium (1% yeast extract, 2% peptone, and 0.01% adenine hemisulfate) with 2% glucose. The cells were then centrifuged and resuspended in synthetic complete medium lacking leucine and tryptophan (SC-Leu-Trp) with 2% raffinose for growing at 30°C with shaking overnight. TALEN expression was induced by culturing cells in YPA medium with 0.002–2% galactose at 30°C for 1–6 h (the induction conditions for each cycle of screening are listed in Supplementary Fig. 2). The cells were then centrifuged and resuspended in SC-Leu-Trp liquid medium with 2% glucose
at 30°C with shaking overnight. On the next day, cells were analyzed on a BD FACS Aria III cell sorting system (BD Biosciences, San Jose, CA) and 20,000–100,000 eGFP-positive cells were collected. After 2 days’ growth at 30°C in SC-Leu-Trp liquid medium with 2% glucose, plasmids from the sorted cells were extracted using Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research, Orange, CA). Sequences encoding the CTS and FokI domain were PCR amplified from the plasmids by Q5 DNA polymerase and gel purified for the next cycle of screening.

A Modified Surrogate Reporter System

A TALEN target site within the sickle human beta-globin (HBB) gene was inserted between the monomeric red fluorescent protein (mRFP) gene and the eGFP gene by overlap extension PCR. The fusion gene was cloned into the pLNCX2 retroviral vector (BD Clontech, Palo Alto, CA) in the presence of the cytomegalovirus (CMV) promoter and stably integrated into the genome of HeLa cells through retroviral transduction according to the manufacturer’s protocol. Transfected cells were selected in 500 μg/mL G418 for 2 weeks. The TALEN genes along with a FLAG tag sequence and a SV40 nuclear localization signal added to the N-terminus were cloned into the pCMV5 mammalian expression vector (Andersson et al., 1989) through the KpnI and SalI sites.

Reporter HeLa cells were routinely maintained in the modified Eagle’s medium (MEM) supplemented with 10% fetal bovine serum (FBS; Hyclone, Logan, UT). Cells were seeded in 24-well plates at a density of 5 × 10⁴ per well. After 24 h, reporter cells were transfected with certain amounts of TALEN expression plasmids using FuGene HD transfection reagent (Promega, Madison, WI) under conditions specified by the manufacturer. For each transfection, the overall plasmid amount was made to be 500 ng by adding the empty pCMV5 plasmid. Cells were trypsinized from their culturing plates 48 h after transfection and resuspended in 300 μL phosphate-buffered saline (PBS) for flow cytometry analysis. Twenty thousand cells were analyzed by a BD LSRII flow cytometer (BD Biosciences) to quantify the eGFP-positive cells.

Western Blot Analysis

HeLa cells in 12-well plates were harvested 24 h after transfection with pCMV5-TALEN-HBB-R plasmid (Sun et al., 2012b). Cells were collected by centrifugation, washed with PBS, and resuspended in 40 μL whole cell lysis buffer (1 M Tris–HCl, pH 6.8, 20% sodium dodecyl sulfate, and 0.1 M dithiothreitol). Proteins were resolved by 4–20% Mini-Protein TGX Precast Gel (Bio-Rad Laboratories, Hercules, CA), transferred onto a nitrocellulose membrane, blocked for 1 h with Tris-buffered saline/0.05% Tween 20 containing 5% nonfat milk, followed by incubation with anti-FLAG tag (1:500) and anti-α-tubulin (1:10,000) antibodies at 4°C overnight. After incubation with anti-mouse horseradish peroxidase-conjugated secondary antibody (1:25,000; Gene-Script, Piscataway, NJ) for 1 h, bands were visualized using SuperSignal West Pico Chemiluminescent Substrate (Pierce, Rockford, IL).

Results

Development of a High-Throughput Screening System

To improve the performance of the existing TALEN technology using directed evolution, we developed a high throughput screening system in S. cerevisiae. This system employs an eGFP-based single-strand annealing reporter construct that is integrated into the yeast genome. The reporter construct contains a divided eGFP gene harboring two AvrXa10 TALE binding sites in a tail-to-tail orientation (Supplementary Table I). The DSB generated by the AvrXa10 TALE homodimer can induce single-strand annealing of the two truncated eGFP DNA fragments and reconstitute a complete and functional eGFP gene (Fig. 1A). Therefore, this system couples the enzymatic DNA cleavage activity with the fluorescent signal of host cells, and enables high-throughput screening of TALEN mutants with improved genome editing.
efficacy. The TALEN gene expression is under the control of a galactose-inducible promoter, thus the selection pressure of directed evolution can be adjusted by galactose concentration in the growth medium and incubation time during induction (Fig. 1B).

**Directed Evolution of TALENs With Improved Activity**

Using a TALEN scaffold bearing a 207 aa NTS and a 63 aa CTS (hereinafter referred to as WT) (Sun et al., 2012b) as a template, we randomly mutagenized the sequences encoding the CTS and the FokI nuclease domain using error-prone PCR and constructed a library of mutants in vivo by gap repair homologous recombination in yeast (Supplementary Fig. 1). TALEN expression was induced by galactose and the eGFP-positive cells were collected by fluorescence-activated cell sorting (FACS). The sorted cells were pooled together and plasmids from the pool were isolated. The regions encoding the CTS and FokI nuclease domain of TALEN genes were PCR-amplified from the plasmids and subjected to another round of screening. This process was continued in an iterative fashion until the TALEN mutants with improved genome editing activity were enriched and identified (Fig. 2). We gradually increased the selection stringency by decreasing the galactose concentration and induction time during the enrichment (Supplementary Fig. 2). After two rounds of mutagenesis and evolution, 14 TALEN mutants with improved activity were generated (Table I, Fig. 3 and Supplementary Fig. 3). The most active mutant (2S3C3) has greater than sevenfold improvement in yeast genome modification compared to WT, whereas GoldyTALEN and the TALEN bearing a Sharkey domain (a FokI domain variant with S35P and K58E mutations) (Guo et al., 2010) only exhibit <3.5-fold increase compared to WT.

**Table I. Information and in vivo activities of TALEN variants.**

<table>
<thead>
<tr>
<th>Label</th>
<th>Mutation</th>
<th>Relative activity (fold of WT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yeast</td>
</tr>
<tr>
<td>WT</td>
<td>—</td>
<td>1.00 ± 0.11</td>
</tr>
<tr>
<td>Sharkey</td>
<td>S35P, K58E</td>
<td>2.38 ± 0.23</td>
</tr>
<tr>
<td>Goldy</td>
<td>L(-33)M, H(-16)K, P(-11)G</td>
<td>3.24 ± 0.15</td>
</tr>
<tr>
<td>S4C5</td>
<td>P(-11)T, H(-6)Q, V521</td>
<td>3.49 ± 0.14</td>
</tr>
<tr>
<td>S4C9</td>
<td>P(-11)H, L182L, R186R</td>
<td>2.90 ± 0.32</td>
</tr>
<tr>
<td>S4C10</td>
<td>K58R, N144K</td>
<td>1.30 ± 0.10</td>
</tr>
<tr>
<td>S7C1</td>
<td>P(-11)R, I82I, H115H, N153K, L168L, L182L</td>
<td>4.40 ± 0.16</td>
</tr>
<tr>
<td>S7C6</td>
<td>P(-11)H</td>
<td>4.11 ± 0.23</td>
</tr>
<tr>
<td>S7C7</td>
<td>S55, I175I, K176Q, L182L</td>
<td>2.83 ± 0.17</td>
</tr>
<tr>
<td>S7C10</td>
<td>K58E, I170L, L182L, E193V</td>
<td>1.76 ± 0.22</td>
</tr>
<tr>
<td>Sunny</td>
<td>P(-11)H, S35P, K58E</td>
<td>4.13 ± 0.27</td>
</tr>
<tr>
<td>2S3C3</td>
<td>P(-11)H, S35P, K58E</td>
<td>7.26 ± 0.38</td>
</tr>
<tr>
<td>2S3C11</td>
<td>P(-11)H, S35P, K58E, A87A, F134Y</td>
<td>6.85 ± 0.42</td>
</tr>
<tr>
<td>2S3C12</td>
<td>L(-33)M, P(-11)H, L181L, S35P, K58E, K123N</td>
<td>4.91 ± 0.40</td>
</tr>
<tr>
<td>2S3C18</td>
<td>P(-11)H, E(-10)K, L7L, S35P, K58E, 170M, P76L, V81V</td>
<td>6.20 ± 0.57</td>
</tr>
<tr>
<td>2S6C1</td>
<td>P(-11)H, N34D, K58E</td>
<td>3.88 ± 0.56</td>
</tr>
<tr>
<td>2S8C6</td>
<td>P(-11)H, E(-10)K, S35P, K58E</td>
<td>5.67 ± 0.47</td>
</tr>
</tbody>
</table>

*The residue numbering is indicated in Supplementary Figure 1B. Silent mutations are shown in gray.

*The activity of WT was normalized to 1. Values represent the average (±standard deviation) of at least three independent experiments.

*Activities in yeast were measured when the TALEN expression was induced in the culture medium containing 0.02% galactose for 4 h.

*Activities in human cells were measured in the modified surrogate reporter system (Fig. 4A) in a dose-limiting condition (32 ng transfected TALEN expression plasmids per 10⁶ cells).

*GoldyTALEN has a shorter and mutated NTS compared with all the other TALENs. Complete sequence is shown in Supplementary Figure 11.
Isolation and Characterization of the SunnyTALEN Scaffold

To test the efficacy of the isolated TALEN mutants in human cells, we modified a surrogate reporter system (Kim et al., 2011) to rapidly and quantitatively gauge the potential of TALEN-driven genome editing (Fig. 4A). The reporter construct encoding a mRFP-eGFP fusion protein was stably integrated into the genome of Hela cells. A TALEN target site within the HBBS gene locus (Supplementary Table I) (Sun et al., 2012b) was inserted between the mRFP and eGFP genes, making the eGFP gene out of frame. In response to the TALEN-induced DSBs, certain insertions and deletions (indels) caused by NHEJ-mediated mutagenesis will make the eGFP gene in frame and restore eGFP function. Each TALEN mutant was under the control of a CMV promoter, and genome editing activity was measured by flow cytometry following transient expression. Although the TALEN bearing a Sharkey domain and GoldyTALEN showed similar or lower genome editing efficacy compared to WT, one TALEN mutant (referred to as SunnyTALEN) has >2.5-fold improvement compared to WT (Table I, Fig. 3 and Supplementary Fig. 4). SunnyTALEN contains a single P(-11)H mutation on the CTS and a Sharkey domain. We performed titrations of the expression plasmids of different TALEN architectures at various amounts and obtained a dose–response curve (Fig. 4B). SunnyTALEN consistently yielded more eGFP-positive cells compared to the other TALEN variants. The P(-11)H mutation of SunnyTALEN serves as a major contributor to its improved efficacy without changing steady state level of protein. The S35P and K58E mutations on the Sharkey domain increase activity marginally, which can be explained by the increased steady state level of protein (Figs. 4B and 5). The lower protein level of GoldyTALEN is consistent with its lower efficacy in human cells than WT. In addition, cells transfected with TALEN variants were subjected to the immunofluorescent analysis for cellular localization. The results showed that all TALEN variants, which contain an N-terminal nuclear localization signal, were predominantly localized in the nuclei (Supplementary Fig. 5).

Application of the SunnyTALEN Scaffold for Human Genome Editing

To further explore the general applicability of the SunnyTALEN scaffold in modifying human genomes, we generated TALENs against 12 additional genomic loci of human embryonic kidney (HEK)293 cells. Using the same TALE DNA binding domains in the SunnyTALEN scaffold, we observed increased rate of gene modification at all the targeted loci compared to WT, with the increased indels ranging from 4% to 33% in response to NHEJ-mediated mutagenesis (Fig. 6A and Supplementary Fig. 6). The SunnyTALEN scaffold also showed higher rate of gene modification compared to the WT TALEN with the Sharkey mutations or with the P(-11)H substitution (Supplementary Fig. 7). Sequencing analysis of the two targeted loci indicated...
that SunnyTALENs induced more small deletions than insertions (Supplementary Fig. 8), with the mutation signatures similar to the first-generation TALENs (Kim et al., 2013b) and GoldyTALENs (Bedell et al., 2012). We also explored whether the rates of TALEN-induced homologous recombination could be enhanced using an eGFP gene conversion assay (Supplementary Fig. 9A) (Sun et al., 2012b). In the linear range of the dose-response curve, we observed a >2-fold increase in homology-directed gene repair using the SunnyTALEN scaffold over WT (Supplementary Fig. 9B), which paralleled the increase observed in NHEJ-mediated gene disruption levels (Fig. 4B).

To prevent the formation of unexpected cleavage-competent homodimers, obligate heterodimeric FokI domains have been used in TALENs to reduce off-target effects and cytotoxicities (Cade et al., 2012; Huang et al., 2011). To investigate the compatibility of the SunnyTALEN scaffold with the obligate heterodimeric FokI domains, we introduced the ELD:KKR FokI pair (ELD denotes Q103E, I116L, and N113D mutations; KKR denotes E107K, I155K, and H154R mutations) (Doyon et al., 2011) that preferentially heterodimerize into the SunnyTALEN scaffold and compared the genome editing efficacy with the WT counterpart in the modified surrogate reporter system (Fig. 4A). The pairwise analysis of these architectures showed that the TALEN

![Figure 5](image)

**Figure 5.** Western blot analysis of TALEN variants in human cells. All the TALEN variants contained a FLAG tag at the N-terminus. a-Tubulin was used as a loading control. An asterisk indicates non-specific bands.

![Figure 6](image)

**Figure 6.** Characteristics of the SunnyTALEN platform in human genome editing. A: The SunnyTALEN scaffold improved NHEJ-mediated gene disruptions at 12 genomic loci in HEK293 cells. Representative gel images of the SURVEYOR nuclease assays are shown in Supplementary Figure 6. B: The SunnyTALEN scaffold is compatible with the obligate heterodimeric FokI nuclease domains. Activities were measured in the modified surrogate reporter system (Fig. 4A) in a dose-limiting condition (32 ng transfected TALEN expression plasmids per 10⁵ cells). Error bars indicate standard deviation of three independent experiments. Empty vector pCMV5 served as a negative control. C: Nuclease-associated cytotoxicities of TALENs. Low dose indicates 32-ng TALEN expression plasmids were transfected per 10⁵ cells. High dose indicates 256-ng TALEN expression plasmids were transfected per 10⁵ cells. A reported toxic zinc finger nuclease (GZF3N) served as a positive control (Szczep et al., 2007). Error bars indicate standard deviation of three independent experiments.
heterodimers with SunnyTALEN scaffold had >4.4-fold improvement in stimulating mutagenesis compared to the WT TALEN heterodimers (Fig. 6B). None of the ELD:ELD or KKR:KKR TALEN dimers were observed to be cleavage competent. To assess the nuclease-associated toxicity, we monitored the phosphorylation of histone H2AX (γ-H2AX), which rapidly responds to chromosomal DSBs and can be used to measure genome-wide DNA damage levels. Although the SunnyTALEN scaffold slightly increased γ-H2AX level over WT, the introduction of the obligate heterodimeric FokI domains reduced γ-H2AX staining to the background level (Fig. 6C). Based on the prediction by TAL Effector Nucleotide Targeter 2.0 (Doyle et al., 2012), a highly similar sequence within the HBD gene that shares only four mismatches compared with the designed TALEN binding site was identified (Supplementary Fig. 10A). Modest off-target cleavage at this site was observed using the SunnyTALEN scaffold at high dose (Supplementary Fig. 10B).

Discussion

In this study, we developed a powerful yeast-based screening system to improve TALEN activity. This system couples enzymatic DNA cleavage with GFP signal of yeast cells and enables directed evolution of TALENs with improved genome editing efficacy. The GFP-positive cells can be isolated by FACS, which enables high-throughput screening of large libraries of mutants with single cell resolution and low false-positive rate. Although applied here in enhancing TALEN activity, this system can be easily adapted to engineer other rare-cutting DNA endonucleases such as homing endonucleases (or meganucleases) (Silva et al., 2011), zinc finger nucleases (ZFNs) (Carroll, 2011), and newly developed clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system (Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2013; Mali et al., 2013), for the purpose of increasing catalytic activities or altering DNA recognition specificities.

In the SunnyTALEN scaffold, the P(-11)H mutation on the CTS plays a major role in activity improvement. Because structural information is not available to visualize that residue, we have built structural models of WT and SunnyTALEN scaffolds with 63 aa CTSs to study the molecular mechanism. Structural models indicate the substitution of a histidine for a proline introduces an additional hydrogen bond between the CTS of the SunnyTALEN scaffold and the DNA backbone, which might increase catalytic activity by positioning the FokI nuclease domain closer to the DNA substrate or to the FokI domain of the other TALEN monomer (Fig. 7). More detailed structural studies will be needed to confirm this hypothesis. In addition to the isolation of a high-efficiency TALEN system, our directed evolution endeavor identified “hot spot” residues (e.g., P(-11), E(-10), N34, S35, K51, K58, F134, and K176) that are critical for the catalytic activity or protein solubility of TALENs (Table I), which provides an important insight for the further improvement of TALEN system by rational protein design or saturation mutagenesis.

Using γ-H2AX as a marker of chromosomal DNA damage, we measured the nuclease-associated genotoxicities in human cells. Even though the increased genome editing efficacy of the SunnyTALEN scaffold is accompanied by increased DNA damage, γ-H2AX level can be reduced by lowering the TALEN dose or introducing obligate
heterodimeric *FokI* nuclease domains (Fig. 6C). Because of the high efficiency of the SunnyTALEN scaffold, it can generate sufficient genetic modifications at dose-limiting conditions without causing detectable genotoxities (Figs. 4B and 6C). Off-target cleavage is dependent on protein concentration as demonstrated for ZFNs (Pattanayak et al., 2011). Here we also showed that lowering the dose of SunnyTALENs decreased the cleavage at an off-target site (Supplementary Fig. 10B). Therefore, it is suggested to optimize the dose of SunnyTALENs for different human cell lines to achieve efficient genome editing with low toxicity. To avoid the off-target events generated by homodimeric ZFNs, obligate heterodimeric mutations were introduced at the dimer interface of the *FokI* cleavage domain to prevent homodimerization based on electrostatic and hydrophobic interactions (Doyon et al., 2011; Miller et al., 2007; Szczepik et al., 2007). Similar principle was applied to TALENs and successfully reduced off-target cleavage and relieved toxicity (Cade et al., 2012). We show here that the incorporation of obligate heterodimeric *FokI* domains completely abolished the formation of cleavage-competent SunnyTALEN homodimers, resulting in reduced genotoxicity (Fig. 6B and C).

In conclusion, we isolated the high-efficiency SunnyTALEN scaffold using a directed evolution strategy. The SunnyTALEN scaffold is portable to many TALE DNA binding domains and compatible with heterodimer TALEN architectures. Although deployed here in yeast and human cells, the SunnyTALEN scaffold could be effective in various organisms and cell types. In addition to the GoldyTALEN scaffold that enables efficient genome modifications in zebrafish and livestock, the SunnyTALEN scaffold serves as a novel second-generation TALEN system and will aid in a variety of applications in medicine and biology.

This work was supported by the Centennial Chair Professorship (H.Z.) in the Department of Chemical and Biomedical Engineering at the University of Illinois at Urbana-Champaign. We thank the University of Illinois Flow Cytometry Facility for assistance with flow cytometry and FACS.

**References**


Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web-site.