Minireview

Engineering and Application of Zinc Finger Proteins and TALEs for Biomedical Research

Moon-Soo Kim[1],* and Anu Ganesh Kini[1]

[1]Department of Chemistry, Western Kentucky University, 1906 College Heights Blvd., Bowling Green, KY 42101, USA
*Correspondence: moon-soo.kim@wku.edu
http://dx.doi.org/10.14348/molcells.2017.0139
www.molcells.org

Engineered DNA-binding domains provide a powerful technology for numerous biomedical studies due to their ability to recognize specific DNA sequences. Zinc fingers (ZF) are one of the most common DNA-binding domains and have been extensively studied for a variety of applications, such as gene regulation, genome engineering and diagnostics. Another novel DNA-binding domain known as a transcriptional activator-like effector (TALE) has been more recently discovered, which has a previously undescribed DNA-binding mode. Due to their modular architecture and flexibility, TALEs have been rapidly developed into artificial gene targeting reagents. Here, we describe the methods used to design these DNA-binding proteins and their key applications in biomedical research.

Keywords: biomedical application, sequence-specific DNA detection, transcriptional activator-like effector, zinc fingers

INTRODUCTION

The Cys2-His2 (C2H2) domain is the most common type of DNA-binding motif found in eukaryotes. The C2H2 ZF domain contains multiple cysteine and histidine residues, which are the most common ligands for the zinc ion in proteins since they use zinc coordination to stabilize their folds (Segal and Meckler, 2013). The DNA-binding activity of ZF domains has been extensively studied, and a number of studies have been conducted to create ZF proteins (ZFPs) that recognize any desired DNA sequence to provide useful new tools for numerous biomedical research applications such as gene regulation, genome engineering and diagnostics.

A new class of DNA binding domain has been recently discovered, which is called a transcriptional activator-like effector (TALE). The recent discovery of TALEs has enabled many scientists to exploit an alternative platform for engineering DNA-binding proteins. TALEs are naturally secreted proteins from plant pathogenic bacteria of the genus *Xanthomonas*. They contain modular DNA-binding domains composed of 33-35 amino acid repeat arrays, where each repeat domain specifies a single DNA base. TALEs have been deployed in DNA targeting for many applications, such as genome engineering and gene regulation (Miller et al., 2011). Numerous studies have shown considerable progress in understanding DNA recognition of ZFPs and TALEs and in developing design methods that could open up more widespread applications for a number of biomedical research opportunities.

ZINC FINGER PROTEINS – STRUCTURE AND DNA RECOGNITION

The C2H2 ZF domain is the most common type of ZF and is one of the most abundantly expressed proteins in eukaryotic cells. ZFs are small, functional and independently folded domains coordinated with zinc molecules in their structure. The C2H2 ZF folds into a compact ββα structure, which is stabilized by zinc coordination and by the conserved hydrophobic core. This ββα framework provides an insight into how ZFPs interact with DNA. Twenty-five of the thirty amino acids in the repeat folds around the zinc to form a 'finger' and the rest of the five amino acids (TGEK(R)P) provide a
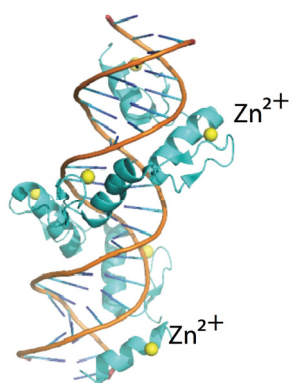
Fig. 1. DNA bound structure of a zinc finger protein (ZFP). ZFP Aart (PDB ID: 2I13) is flanking the major groove of DNA, making contact with the edge of the nucleotide bases. Aart is a designed six finger protein with pentapeptide linkers, which recognizes an A-rich 18 bp sequence (Segal et al., 2006). The DNA is slightly unwound in order to allow 11.3 bp/turn.

short consensus linker between consecutive fingers (Moore et al., 2001). The zinc ion is tetrahedrally coordinated between two cysteine and two histidine residues, which stabilizes the fingers.

Amino acids in each ZF have affinity towards specific nucleotides, causing each finger to selectively recognize 3-4 nucleotides of DNA. Multiple ZFs can be arranged into a tandem array and recognize a set of nucleotides on the DNA. It is possible to create modules of six ZFs that can potentially recognize 18 bp of DNA, which would be sufficient enough to recognize a unique DNA sequence in the human genome. The α-helix of each finger fits into the major groove of the DNA, causing the protein to wrap around the DNA, as shown in the crystal structure of Aart (Fig. 1) (Segal et al., 2006). Three amino acid residues at positions -1, 3, and 6 on the α-helix make contacts with the 3', middle, and 5' nucleotides, respectively. In addition, amino acids at positions -2, 1, and 5 can make direct or water-mediated contacts to the phosphate backbone of the DNA (Segal et al., 2003). The amino acid at position 2 is also involved in the contact with other helix residues. Different ZFPs of various lengths can be generated, which allow for recognition of almost any desired DNA sequence out of the possible 64 triplet subsites (Dreier et al., 2001).

ENGINEERING OF ZINC FINGER PROTEINS (ZFPs)

Modular assembly

Multiple approaches have been taken to identify optimized individual ZF modules to recognize one of the 64 possible 3 bp DNA subsites using a combination of rational design and selection (Beerli and Barbas, 2002; Dreier et al., 2000; 2005; Segal et al., 1999; Wu et al., 1995). Phage display libraries were constructed and selected, which contained all amino acid residues randomized in the α-helix of the ZF (Beerli and Barbas, 2002; Wu et al., 1995). By selecting for phage using oligonucleotides that contain a specific 3 bp subsite, ZF recognition modules that bind to specific 3 bp subsites were

isolated. In principle, multi-finger proteins can be constructed by assembling predefined ZF modules in any order to recognize any desired DNA sequence in a modular fashion, which is referred to as modular assembly (Bhakta and Segal, 2010). A set of modular assembly fingers described here was developed by Barbas and his colleagues. Another set of modular assembly fingers was developed by ToolGen, which was based on selections of ZFs occurring naturally in the human genome as opposed to synthetic variants of Zif268 (Bae et al., 2003). The Barbas and ToolGen domains are the two most commonly used sets of modular assembly fingers. Both domains cover all 3 bp GNN, most ANN, many CNN and some TNN triplets (where N can be any of the four nucleotides). Both have a different set of fingers, which allows for searching and coding different ZF modules as needed (Bhakta and Segal, 2010; Gersbach et al., 2014). The main advantage of this approach is that ZFs can be assembled in any order and no selection step is required. Many of the engineered ZF domains that were constructed via modular assembly were shown to have higher specificity as compared to naturally occurring ZF domains by greater than 100-fold in the three finger context (Gersbach et al., 2014). Since modular assembly is one of the most popular methods for constructing ZFPs, it has been widely used in numerous applications, such as nucleases, transposases, recombinases, integrases, and gene regulators (Camenisch et al., 2008; Gordley et al., 2009; Kolb et al., 2005).

Oligomerized pool engineering (OPEN)

To minimize context-dependent effects of modular assembly involving the position of a finger in the protein and the sequence of neighboring fingers, a combinatorial selection-based oligomerized pool engineering (OPEN) strategy was developed by J. Keith Joung for constructing multi-finger arrays (Maeder et al., 2009). Before the OPEN approach, the Pabo group developed the bacterial two-hybrid (B2H) system, which involves a two-step process (Joung et al., 2000). The first step is performed to enrich a finger with out-of-context 3 bp subsites as in modular assembly. Then, a second round of selection enriches ZFs in the context of the full, intended DNA target sequence. Joung and colleagues developed OPEN as an optimized version of this method. An archive of pre-selected ZF pools is used in OPEN, each consisting of a maximum of 95 different fingers targeted to a specific 3 bp subsite at a defined position (Maeder et al., 2008). Appropriate finger pools from the archive are recombined to create a small library of multi-finger arrays for a target 9 bp site of interest. Members of this library are then screened using the B2H selection system, where ZF binding to its designed site activates the expression of selectable marker genes. The efficiency and success rate of this method is approximately 70-80% for obtaining ZF arrays capable of activating transcription in B2H strains (Maeder et al., 2009). OPEN ZF arrays and zinc finger nucleases (ZFNs) are publicly available from the Zinc Finger Consortium Database (Maeder et al., 2009). Although OPEN has sequence constraints in their ZFNs, it has been used to develop successful ZFNs targeting sites in human cells and plants (Curtin et al., 2011; Sebastiano et al., 2011).

APPLICATION OF ZINC FINGER PROTEINS IN BIOMEDICAL RESEARCH

Zinc finger nucleases (ZFNs)

Upon linking ZFs to a nuclease domain, ZFNs are constructed to recognize and cleave DNA at a desired location. The cleavage domain from the type IIs restriction enzyme FokI is fused to ZFs, thereby creating a DNA double-strand break (DSB) at targeted sites. FokI domains must dimerize to cleave DNA. Hence, two ZFs are fused with two FokI cleavage domains to assemble functional ZFNs. Two ZF-FokI monomers bind independently in an inverted tail-to-tail orientation and with a 5-7 bp spacer sequence recognized by the cleavage domain between the binding sites (Gaj et al., 2013a) (Fig. 2).

A ZFN-induced DSB will stimulate cellular DNA repair mechanisms either by error-prone non-homologous end joining (NHEJ) or precise homologous recombination (HR). NHEJ-mediated repair often results in small insertion or deletion (indel) errors at the targeted site. HR involves a precise addition of an exogenous nucleotide sequence that is complementary to the sequence on the broken double-stranded DNA, making it easy to incorporate any nucleotide sequence of choice into the DNA.

Gene editing mediated by ZFNs has been applied to correct disease-causing mutations associated with sickle cell disease (Sebastiano et al., 2011; Zou et al., 2011), α_1-antitrypsin deficiency (Yusa et al., 2011), hemophilia B (Li et al., 2011a) and Parkinson's disease (Soldner et al., 2011). For the correction of Parkinson's disease-associated mutations, ZFN-mediated genome editing was combined with induced pluripotent stem cells (iPSCs) technology (Soldner et al., 2011). This approach enables the genetic correction of point mutations in the α-synuclein gene in patient-derived human iPSCs. Another example of combination of ZFNs with iPSCs technology was also found in the correction of the E6V mutation in the β-globulin gene for sickle cell disease (Sebastiano et al., 2011).

ZFN-mediated gene disruption has been taken to clinical trials (NCT00842634 and NCT01044654) for treating HIV (Holt et al., 2010; Perez et al., 2008). ZFNs have been used to confer HIV-1 resistance in CD4[+] T cells by disrupting the co-receptor chemokine (C-C motif) receptor type 5 (CCR5). This ZFN approach potentially results in a heritable gene kno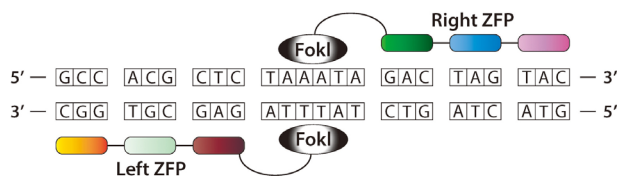ckout of CCR5 and consequently HIV resistance (Urnov et al., 2010). This could allow ZFN-modified CD4[+] T cells to potentially reconstitute immune function in patients with HIV/AIDs by maintaining an HIV-resistant CD4[+] T cell population (Urnov et al., 2010). ZFNs were also used to engineer hematopoietic stem and progenitor cells (HSPCs) so that the stem cells could directly mutate and become resistant to HIV (Li et al., 2013).

Zinc finger recombinases (ZFRs)

Site-specific recombinases (SSRs) are capable of recognizing 30-40 bp sequences and catalyzing excision, inversion, or integration between defined segments of DNA (Grindley et al., 2006). Due to their strict target specificity it poses a challenge to use SSRs in cells and organisms that have artificially induced recombination sites or pre-existing recombination sites (Gaj et al., 2013b). To address this challenge, ZFRs were introduced as an effective alternative to the conventional site-specific recognition systems (Gordley et al., 2007). ZFRs catalyze recombination between specific ZF target sites. Like ZFNs, they consist of two inverted ZFs on double-stranded DNA with the recombinase domain exercising its catalytic activity in the 20 bp central flanking region (Smith and Thorpe, 2002). Successful re-engineering of serine recombinases was studied to explore the specificity and effectiveness of ZFRs. Using this approach, Gosh. et al generated enhanced hybrid recombinases based on activated catalytic domains derived from the resolvase/invertase family of serine recombinases (Gin, Hin, Tn3 and $\gamma\delta$) (Gaj et al., 2014b). Through rational design and directed evolution, they re-engineered the serine recombinase dimerization interface. The re-engineered hybrid recombinases showed higher specificity with low toxicity, indicating the potential of these enzymes in a wide range of applications for genome engineering and gene therapy (Gaj et al., 2013b).

Zinc finger transposases

The Sleeping Beauty (SB) transposon is an integrating vector system capable of inserting expression cassettes with high stability. However, the SB insertion profile is close-to-random in the genome, and random genomic insertion can cause unwanted mutagenesis of endogenous genes (Voigt et al., 2012). To address this problem, an attempt for targeted transposon insertion has been made such that the transposase or the transposon vector DNA is physically linked to a DNA-binding domain (DBD). In this way, the transposase/transposon complex is tethered to defined sites in the genome and is able to facilitate integration of the transposon into the adjacent intended DNA (Voigt et al., 2012). Fusion of a ZF with SB transposase resulted in a fourfold enrichment of the transposon insertion as compared to native SB transposase (Voigt et al., 2012). In another study, Zif268 was fused to the C-terminus of ISY100 transposase, resulting in highly specific integration into TA dinucleotides positioned 6-17 bp to one side of a binding site for Zif268 (Feng et al., 2010). The classical Gal4 DBD was also fused to a piggyback (PB) transposase to bias genomic insertion to specific sites, upstream activating sequence (UAS) Gal4 recognition sites (Owens et al., 2012). Gal4-PB fusion proteins were able to target transposition near to UAS sites,

Fig. 2. Schematic diagram of a zinc finger nuclease (ZFN) dimer bound to target DNA. ZF arrays are fused to the FokI nuclease domain to make a custom nuclease that can recognize unique left and right half-sites. The two ZFNs must bind in an inverted tail-to-tail orientation with their C-termini facing each other. The optimal spacing between the half-sties is 5-7 bp.

which were randomly integrated throughout the genome, as compared to native PB transposase.

Zinc finger-artificial transcription factor (ATF)

DNA-binding domains including ZFs can be engineered to regulate expression of specific genes by fusing them to transcriptional or epigenetic effector domains, thus generating artificial transcription factors (ATFs). In principle, ATFs are comprised of a DNA-binding domain, a transcriptional activator (VP16 and p65 domains) or repressor (KRAB domain), and a nuclear localization signal (NLS) to ensure the efficient transport of ATFs into the nucleus. Engineered ZFPs were fused to VP64 and KRAB domains to create synthetic activators and repressors, respectively (Beerli et al., 1998). These ATPs were demonstrated to up- and down-regulate the endogenous *ERBB2* and *ERBB3* genes in human cells (Beerli et al., 2000). Later, engineered ZF repressors were shown to down-regulate the HIV promotor and reduce HIV replication in primary cells up to 100-fold (Segal et al., 2004). As a therapeutic approach to sickle cell disease and β-thalassemia, ZF ATFs were designed to target the promotor region of the γ-globin gene, resulting in up-regulation of γ-globin expression in human cell lines (Graslund et al., 2005) and activation of the silent γ-globin gene in primary human hematopoietic stem cells in an *in vivo* mouse model (Wilber et al., 2010). ZF ATFs have been designed to reactivate the paternal UBE3A gene that is silenced by imprinting in an attempt to develop a molecular therapy for Angleman syndrome (Bailus and Segal, 2014).

Protein delivery

ZFNs are intrinsically cell permeable (Gaj et al., 2012), which is attributed to the net positive charge of ZF domains (Gaj et al., 2014a). Cell penetrating ZF domains were successfully proven to be good protein transduction reagents (Gaj et al., 2014a). Gaj et al. demonstrated that when the N-terminus of firefly luciferase was genetically fused to two or three fingers, it resulted in cell penetrating properties as effective as Lipofectamine-mediated plasmid transfection. These protein-fused ZFs are capable of delivering functional proteins into primary and transformed mammalian cells (Gaj et al., 2014a). This study also showed that ZFPs enter the cells mainly through macropinocytosis and at low frequencies through caveolin-dependent endocytosis.

DNA diagnostics

A system called Sequence Enabled Reassembly of β-lactamase (SEER-LAC) consists of two split enzymatic domains of β-lactamase that would reassemble into a full-length enzyme upon ZFPs binding to their target DNA (Ooi et al., 2006). Kim et al. (2011) developed a ZFP array combined with the SEER-LAC system for DNA diagnostic applications. The ZFP array with the SEER-LAC system generated DNA-dose dependent signals with a visual readout and allowed for a quantitative assay. Their result suggested the potential use of this system to develop a point-of-care (POC) diagnostic for pathogen detection.

Cytosine on CpG is frequently methylated in certain genes, causing epigenetic silencing. Detection of DNA methylation is an excellent diagnostic tool for early detection of different carcinomas and adenoma. Ghosh et al. (2006) used the SEER-LAC system fused with an engineered ZFP and a methyl binding domain for direct detection of methylated dsDNA.

TRANSCRIPTIONAL ACTIVATOR-LIKE EFFECTORS (TALEs) – STRUCTURE AND DNA RECOGNITION

TALE represents the largest effector family and functions in transcriptional activation of plant genes. Their unique structure encompasses a DNA-binding region that enables TALEs to bind specifically to the promoter region on DNA. Their binding specificity can be predictable since two hypervariable amino acids within the repeat domain known as repeat variable di-residues (RVDs) determine the nucleotide to which the particular repeat binds (Boch et al., 2009).

The repetitive nature of TALE DNA-binding domains led to the binding code being deciphered in 2009 (Boch et al., 2009). As shown in Fig. 3, amino acids at positions 12 and 13 within a 34-amino acid repeat are called RVDs, which direct nucleotide specificity on the target DNA. The tandem polymorphic amino acid repeats of TALEs are located in the central DNA-binding region. Each RVD recognizes a single DNA base and different RVDs have variable affinity for different nucleotides. The four most common RVDs are HD, NG, NI, and NN, specifying C, T, A, and G, respectively (Boch et al., 2009). There are about 24 known unique RVDs with seven of the most common being HD, NG, HG, NN, NS, NI and N∗ (N∗ corresponds to a 33 amino acid repeat with a missing residue within the RVD loop) (Mak et al., 2013). Thus, the number of repeats including the last truncated repeat and the series of RVDs determine the length and the nucleotide composition of the target that they would recognize. TALEs flank the major groove of the DNA helix (Deng et al., 2012; Mak et al., 2012) with the RVDs making contact with the DNA target as shown in Fig. 4. One of these structures is the PthXho1 bound to its target DNA, which shows the presence of two α-helices connected by a loop of RVDs
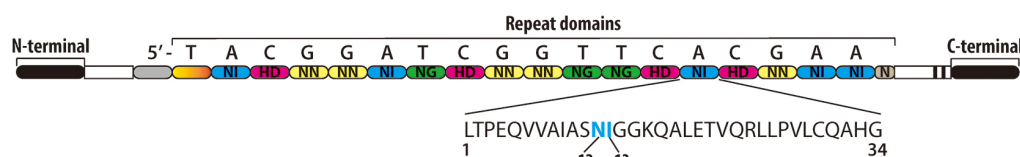

Fig. 3. DNA binding recognition of TALEs. A central DNA-binding region of TALEs contains an array of multiple repeats that are almost identical except for two amino acids at positions 12 and 13 termed repeat variable diresidues (RVDs). Each RVD specifies one DNA base.
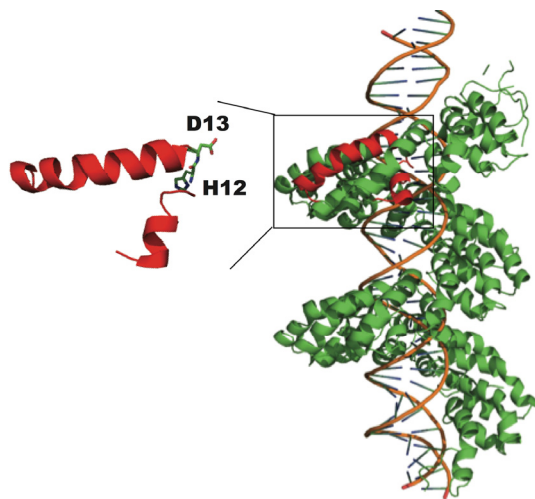
Fig. 4. Crystal structure of PthXo1 bound to DNA. PthXo1 contains 23.5 repeats. The figure shows PthXo1 making contact with a 36 bp dsDNA and the HD RVDs at the 12th and 13th position within the repeat that recognize a single DNA base (PDB ID: 3UGM) (Mak et al., 2012).

that makes contact with the DNA. The target sequence of all naturally occurring TALEs begins with a thymine (T) nucleotide at the 5' end, which is important for the functionality of the TALE's activity (Boch et al., 2009). By deciphering the TALE RVD code, it has been possible to program TALEs with high target specificity and selectivity (Moscou and Bogdanove, 2009). However, a very remarkable study (Rogers et al., 2015) was carried out recently using 21 TALE proteins of different lengths containing all possible consecutive pairs of repeats to identify the influence of these repeats on TALE-DNA binding specificity. Their results infer that not only the affinity of the RVD governs DNA binding, but binding also depends strongly on the base disfavored by the RVD. For example, HD had the highest affinity for C and strongly disfavored G.

CONSTRUCTION OF TALEs

Golden Gate cloning

The major obstacle in the routine usage of TALEs is that the assembly of repeat TALE arrays can be challenging because of extensive identical repeat sequences. To address this issue, methods for achieving rapid assembly of TALE arrays have been studied. One of these is Golden Gate cloning, which allows several DNA fragments to be assembled in a single cloning step (Engler et al., 2008). The assembly may be manipulated to generate sequences of choice without depending on site-specific restriction enzymes since Golden Gate cloning utilizes Type IIs restriction enzymes that cleave outside the recognition sites, creating 4 bp overhangs (sticky ends). The 4 bp overhang can be any four nucleotide sequence of choice, enabling multiple compatible DNA fragments to be ligated together linearly in a single cloning restriction-ligation step.

Morbitzer et al. (2011) developed a rapid, efficient, and low-cost approach for Type IIs enzyme-mediated assembly of repeat modules, which involves fusion of individual TALE repeat modules into a tandem array. This approach allows them to fuse two repeat sub-arrays containing seven and ten repeat-modules into a functional designer TALE (dTALE). dTALE assembly was carried out by two consecutive BsaI cut-ligation steps, followed by BpiI cut-ligation. Their approach resulted in generating a full-length dTALE gene with a high level of sequence fidelity based on sequence-validated plasmids, and not involving PCR.

Another efficient method for assembly of TALE constructs was reported by Cermak et al. (2011) using Golden Gate cloning. The approach allows for assembly of novel repeat arrays for TALE nucleases (TALENs), TALEs, and TALE fusion proteins in just two cloning steps using a set of sequence-verified modules. Golden Gate reaction 1 was performed to build arrays of 1-10 repeats, followed by a Golden Gate reaction 2 to join arrays in a backbone vector to create the final construct TALEN monomer with a 16 RVD array. The software used to design TALENs in their study is available for use as an online tool (http://boglabx.plp.iastate.edu/TALENT/) to identify sequences of recognition sites for the left and right TALEN monomers and the spacer sequence. The identified binding sites are converted into appropriate RVD sequences using the four most commonly used RVDs (NI, HD, NN and NG, specific to A, C, G and T, respectively). Their TALENs were demonstrated to be active in a yeast DNA cleavage assay and effective in gene targeting in human cells and *Arabidopsis*. Similarly, Zhang and his group developed a toolbox to construct different TALE-TFs (TALE transcription factor) and TALENs (TALE nucleases) that can bind to different lengths of DNA (Sanjana et al., 2012).

FLASH

The Fast Ligation-based Automated Solid-phase High-throughput (FLASH) assembly method was developed for rapid construction of large numbers of TALE repeat arrays (Reyon et al., 2012). FLASH allows for high-throughput construction of TALE repeat arrays and ligation of multiple individual TALE repeats in a unidirectional fashion. DNA fragments encoding TALE repeats are assembled on solid-phase magnetic beads, which enables serial restriction digestion reactions, purification and ligation, avoiding the need for column-based washing or purification. The interlaced washing steps between ligation events facilitate the desired order of ligations. The final full-length TALE repeat arrays are released from the beads after a restriction digest, which can be then cloned into a suitable expression vector of choice. One can construct DNA fragments encoding 24 or 96 different TALE repeat arrays in a day using manual or automated FLASH methods, respectively (Reyon et al., 2012). All of the 48 TALEN pairs assembled by FLASH were shown to possess significant EGFP gene disruption activities in a human cell-based assay (Reyon et al., 2012). In addition, FLASH-assembled TALENs were tested for modifying endogenous genes involved in human cancer and epigenetics in human cells. It was found that 84 of the 96 TALENs displayed efficient NHEJ-mediated mutagenesis at the intended target

sites (Reyon et al., 2012).

To facilitate the high-throughput design of FLASH TALE repeat arrays, Joung and his group improved the Zinc Finger and TALE Targeter software (ZiFiT Targeter) (http://ZiFitBeta. partners.org) (Reyon et al., 2013). The upgraded software enables identification of potential target sites and provides all the information needed about the plasmids for the construction of TALE repeats with FLASH.

APPLICATION OF TALEs IN BIOMEDICAL RESEARCH

TALE nucleases (TALENs)

The non-specific FokI nuclease domain can be fused to TALEs to create TALE nucleases (TALENs), which can produce a DNA DSB. FokI cleavage domains as a dimer are attached to the C-terminal end of the two TALEs, with the two TALEs placed tail to tail (Fig. 5). TALENs are designed in pairs to make contact with the two opposing strands of the target DNA, separated by a spacer to provide the FokI nuclease domains with enough space to dimerize and create a DNA DSB, as described in the section on ZFNs. Repair of TALEN-mediated DSBs was shown to create efficient targeted alteration of endogenous genes in several model organisms, including plants (Cermak et al., 2011), yeast (Li et al., 2011b), zebrafish (Sander et al., 2011), human somatic cells (Cermak et al., 2011; Mussolino et al., 2011) and pluripotent stem cells (Hockemeyer et al., 2011).

Sun et al. (2012) have constructed and optimized TALENs from a TALE AvrXa10 by manipulating the N-terminal and C-terminal extensions on either side of the repeat domain along with the spacer length of each effector binding element (EBE). Optimized TALENs showed efficient cleavage of target DNA in the human β-globin gene associated with sickle cell disease with little or no cytotoxicity. Ousterout et al. (2013) have successfully used TALENs to manipulate the nucleotide sequence of the protein dystrophin that is involved in Duchenne Muscular dystrophy disease. Exon 51 was deleted via NHEJ, which corrected the reading frame of the gene and caused successful expression of the protein. The TALEN system was used for HIV-1 gene therapy, resulting in approximately 45% disruption of the CCR5 gene (Mussolino et al., 2011). A similar level of gene disruption was also achieved using ZFNs. However, TALENs showed much lower cytotoxicity with significantly reduced off-target activity as compared to ZFNs.

TALE recombinases

SSRs have emerged as genome engineering tools for manipulating DNA because of their high specificity. To alter the specificity of SSRs, the DBD of SSRs can be replaced by custom-designed DBDs such as ZFs and novel DNA-binding TALEs. In the TALE-recombinase system (TALER architecture), TALEs would bring specificity for inducing a DNA DSB while the recombinase would assure homology directed insertion of exogenous DNA. The first attempt to generate a chimeric TALER was carried out by Barbas's group (Mercer et al., 2012). They created a library of truncated TALE variants to identify optimized TALER fusions with a catalytic domain from the DNA invertase from Gin. Their study showed that TALERs can be used to recombine any DNA sequence in bacteria and mammalian cells, which may overcome the limitation of the modular targeting capacity of ZFRs. They also demonstrated the reprogrammability of the recombinase's catalytic specificity.

TALE transposases

The non-viral PB transposable element fused with the Gal4 DBD has been studied to address the problem of integrating viral vectors associated with insertions at unwanted sites (Owens et al., 2012). One year later, Owens et al. (2013) generated hyperactive PB transposases fused with custom-designed TALEs to target the first intron of the human CCR5 gene. They have demonstrated targeted transposition to the CCR5 genomic safe harbor, which allows for stable expression of a transgene across multiple cell types (Owens et al., 2013).

TALE-artificial transcription factors (TALE-ATFs)

Engineered TALEs can be fused to transcriptional activator and repressor domains to construct artificial transcription factors (ATFs). TALE-ATFs have been successfully used as gene-specific activators and repressors (Maeder et al., 2013; Mahfouz et al., 2012; Perez-Pinera et al., 2013; Zhang et al., 2011). Engineered TALEs fused with a VP64 domain were shown to target a wide spectrum of DNA sequences at a similar or greater level compared to ZF-ATF bearing a VP64 domain (Zhang et al., 2011). Maeder et al. (2013) constructed a large series of TALE activators with various numbers of repeats and tested their activity in stimulating expression of the endogenous human *VEGF-A* gene. In their study, most TALE activators bearing a VP64 domain significantly
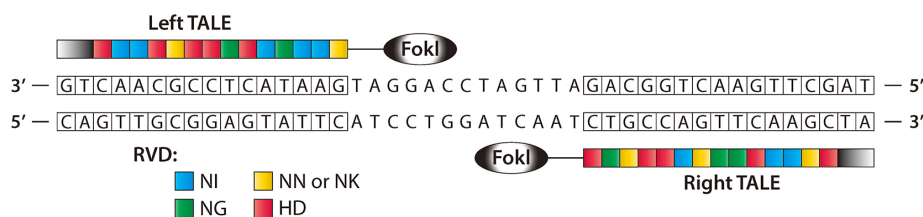

Fig. 5. Schematic diagram of a TALE nuclease (TALEN) dimer bound to target DNA. TALE arrays are fused to the FokI nuclease domain to make a custom nuclease that can recognize unique left and right half-sites. The two TALE binding sites are separated by a spacer of 12-20 bp in length.

induced expression of the *VEGF-A* gene in a very wide range from 5.3- to 114-fold. They also found that both VP64 and p65 TALE activators synergistically induced endogenous gene expression at even higher levels compared to each activator expressed individually (Maeder et al., 2013).

CONCLUSIONS

Engineered ZFPs and TALEs when fused with nucleases, repressors or activators are useful for targeting and manipulating a DNA sequence of interest. ZFPs and ZFNs can be custom-designed depending on the target DNA sequence, but ZFPs have shown a certain sequence preference (5'-GNN-3'). TALEs on the other hand have a flexible and modular structure making it possible to target any desired DNA sequence with robust programmability. Both ZFNs and TALENs have been studied as therapeutic agents with numerous clinical and diagnostic applications as described here.

Distinct from ZFPs and TALEs, clustered regularly interspaced short palindromic repeats (CRISPRs) along with the CRISPR associated proteins (Cas) have recently emerged as an alternative DNA targeting platform. The CRISPR/cas systems depend upon a small database of CRISPR RNAs (crRNA) requiring only programming of a 20-22 bp single-guide RNA (sgRNA) (Jiang and Doudna, 2015). The type II CRISPR/cas system can be engineered as a chimeric 'single-guide RNA' by simply connecting the 3' end of the crRNA to the 5' end of the transactivating crRNAs (tracrRNAs) with a linker sequence (Jiang and Doudna, 2015). This assembly can efficiently direct the Cas9 protein to a target DNA sequence matching the 20 bp RNA guide-sequence and induce a double-strand break in the genome of eukaryotic cells (Jinek et al., 2012). By changing the DNA target sequence within the guide RNA, Cas9 can be retargeted to cleave virtually any DNA sequence in the genome. The simplicity of guide RNA design is an advantage over ZFPs and TALEs since CRISPR/cas technology does not require protein engineering depending on target DNA sequence. Therefore, CRISPR/cas systems can serve as effective tools in genome engineering with greater efficiency and fewer off-target binding events. However, CRISPR/cas technology is still new and future studies are needed to address questions related to DNA-binding specificity in the context of complex genomes.

A wide range of applications have been developed using ZFPs and TALEs so far. ZFPs and TALEs clearly provide a powerful and versatile tool for gene targeting and genome engineering. This being said, there is still considerable potential for researchers to look for new applications for diverse biomedical studies. It will be interesting to see the full potential of CRISPR/cas9 technology for biomedical research and applications.

REFERENCES

Bae, K.H., Kwon, Y.D., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y., Park, J., et al. (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. Nat. Biotechnol. *21*, 275-280.

Bailus, B.J., and Segal, D.J. (2014). The prospect of molecular therapy for Angelman syndrome and other monogenic neurologic disorders. BMC Neurosci. *15*, 76.

Beerli, R.R., and Barbas, C.F., 3rd (2002). Engineering polydactyl zinc-finger transcription factors. Nat. Biotechnol. *20*, 135-141.

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F., 3rd (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. USA *95*, 14628-14633.

Beerli, R.R., Dreier, B., and Barbas, C.F., 3rd (2000). Positive and negative regulation of endogenous genes by designed transcription factors. Proc. Natl. Acad. Sci. USA *97*, 1495-1500.

Bhakta, M.S., and Segal, D.J. (2010). The generation of zinc finger proteins by modular assembly. Methods Mol. Biol. *649*, 3-30.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Camenisch, T.D., Brilliant, M.H., and Segal, D.J. (2008). Critical parameters for genome editing using zinc finger nucleases. Mini Rev. Med. Chem. *8*, 669-676.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. *39*, e82.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., Reyon, D., Dahlborg, E.J., Goodwin, M.J., Coffman, A.P., et al. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol. *156*, 466-473.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science *335*, 720-723.

Dreier, B., Segal, D.J., and Barbas, C.F., 3rd (2000). Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. J. Mol. Biol. *303*, 489-502.

Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D., and Barbas, C.F., 3rd (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. J. Biol. Chem. *276*, 29466-29478.

Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Koksch, B., and Barbas, C.F., 3rd (2005). Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. J. Biol. Chem. *280*, 35588-35597.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

Feng, X., Bednarz, A.L., and Colloms, S.D. (2010). Precise targeted

integration by a chimaeric transposase zinc-finger fusion protein. Nucleic Acids Res. 38, 1204-1216.

Gaj, T., Guo, J., Kato, Y., Sirk, S.J., and Barbas, C.F., 3rd (2012). Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat. Methods 9, 805-807.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013a). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397-405.

Gaj, T., Mercer, A.C., Sirk, S.J., Smith, H.L., and Barbas, C.F., 3rd (2013b). A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic Acids Res. 41, 3937-3946.

Gaj, T., Liu, J., Anderson, K.E., Sirk, S.J., and Barbas, C.F., 3rd (2014a). Protein delivery using Cys2-His2 zinc-finger domains. ACS Chem. Biol. 9, 1662-1667.

Gaj, T., Sirk, S.J., Tingle, R.D., Mercer, A.C., Wallen, M.C., and Barbas, C.F., 3rd (2014b). Enhancing the specificity of recombinase-mediated genome engineering through dimer interface redesign. J. Am. Chem. Soc. 136, 5047-5056.

Gersbach, C.A., Gaj, T., and Barbas, C.F., 3rd (2014). Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. Acc. Chem. Res. 47, 2309-2318.

Ghosh, I., Stains, C.I., Ooi, A.T., and Segal, D.J. (2006). Direct detection of double-stranded DNA: Molecular methods and applications for DNA diagnostics. Mol. Biosyst. 2, 551-560.

Gordley, R.M., Smith, J.D., Graslund, T., and Barbas, C.F., 3rd (2007). Evolution of programmable zinc finger-recombinases with activity in human cells. J. Mol. Biol. 367, 802-813.

Gordley, R.M., Gersbach, C.A., and Barbas, C.F., 3rd (2009). Synthesis of programmable integrases. Proc. Natl. Acad. Sci. USA 106, 5053-5058.

Graslund, T., Li, X., Magnenat, L., Popkov, M., and Barbas, C.F., 3rd (2005). Exploring strategies for the design of artificial transcription factors: targeting sites proximal to known regulatory regions for the induction of gamma-globin expression and the treatment of sickle cell disease. J. Biol. Chem. 280, 3707-3714.

Grindley, N.D., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of site-specific recombination. Annu. Rev. Biochem. 75, 567-605.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. 29, 731-734.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C., et al. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat. Biotechnol. 28, 839-847.

Jiang, F., and Doudna, J.A. (2015). The structural biology of CRISPR-Cas systems. Curr. Opin. Struct. Biol. 30, 100-111.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Joung, J.K., Ramm, E.I., and Pabo, C.O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proc. Natl. Acad. Sci. USA 97, 7382-7387.

Kim, M.S., Stybayeva, G., Lee, J.Y., Revzin, A., and Segal, D.J. (2011). A zinc finger protein array for the visual detection of specific DNA sequences for diagnostic applications. Nucleic Acids Res. 39, e29.

Kolb, A.F., Coates, C.J., Kaminski, J.M., Summers, J.B., Miller, A.D., and Segal, D.J. (2005). Site-directed genome modification: nucleic

acid and protein modules for targeted integration and gene correction. Trends Biotechnol. 23, 399-406.

Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L., et al. (2011a). In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature 475, 217-221.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res. 39, 6315-6325.

Li, L., Krymskaya, L., Wang, J., Henley, J., Rao, A., Cao, L.F., Tran, C.A., Torres-Coronado, M., Gardner, A., Gonzalez, N., et al. (2013). Genomic editing of the HIV-1 coreceptor CCR5 in adult hematopoietic stem and progenitor cells using zinc finger nucleases. Mol. Ther. 21, 1259-1269.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol. Cell 31, 294-301.

Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F., and Joung, J.K. (2009). Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. Nat. Protoc. 4, 1471-1501.

Maeder, M.L., Linder, S.J., Reyon, D., Angstman, J.F., Fu, Y., Sander, J.D., and Joung, J.K. (2013). Robust, synergistic regulation of human gene expression using TALE activators. Nat. Methods 10, 243-245.

Mahfouz, M.M., Li, L., Piatek, M., Fang, X., Mansour, H., Bangarusamy, D.K., and Zhu, J.K. (2012). Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Mol. Biol. 78, 311-321.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 335, 716-719.

Mak, A.N., Bradley, P., Bogdanove, A.J., and Stoddard, B.L. (2013). TAL effectors: function, structure, engineering and applications. Curr. Opin. Struct. Biol. 23, 93-99.

Mercer, A.C., Gaj, T., Fuller, R.P., and Barbas, C.F., 3rd (2012). Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic Acids Res. 40, 11163-11172.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29, 143-148.

Moore, M., Klug, A., and Choo, Y. (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. Proc. Natl. Acad. Sci. USA 98, 1437-1441.

Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res. 39, 5790-5799.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 39, 9283-9293.

Ooi, A.T., Stains, C.I., Ghosh, I., and Segal, D.J. (2006). Sequence-enabled reassembly of beta-lactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. Biochemistry 45, 3620-3625.

Ousterout, D.G., Perez-Pinera, P., Thakore, P.I., Kabadi, A.M., Brown,

M.T., Qin, X., Fedrigo, O., Mouly, V., Tremblay, J.P., and Gersbach, C.A. (2013). Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. Mol. Ther. *21*, 1718-1726.

Owens, J.B., Urschitz, J., Stoytchev, I., Dang, N.C., Stoytcheva, Z., Belcaid, M., Maragathavally, K.J., Coates, C.J., Segal, D.J., and Moisyadi, S. (2012). Chimeric piggyBac transposases for genomic targeting in human cells. Nucleic Acids Res. *40*, 6978-6991.

Owens, J.B., Mauro, D., Stoytchev, I., Bhakta, M.S., Kim, M.S., Segal, D.J., and Moisyadi, S. (2013). Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. Nucleic Acids Res. *41*, 9197-9207.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L., et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. *26*, 808-816.

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat. Methods *10*, 239-242.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. *30*, 460-465.

Reyon, D., Maeder, M.L., Khayter, C., Tsai, S.Q., Foley, J.E., Sander, J.D., and Joung, J.K. (2013). Engineering customized TALE nucleases (TALENs) and TALE transcription factors by fast ligation-based automatable solid-phase high-throughput (FLASH) assembly. Curr. Protoc. Mol. Biol. *Chapter 12*, Unit 12 16.

Rogers, J.M., Barrera, L.A., Reyon, D., Sander, J.D., Kellis, M., Joung, J.K., and Bulyk, M.L. (2015). Context influences on TALE-DNA binding revealed by quantitative profiling. Nat. Commun. *6*, 7440.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat. Biotechnol. *29*, 697-698.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. Nat. Protoc. *7*, 171-192.

Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F., et al. (2011). In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. Stem Cells *29*, 1717-1726.

Segal, D.J., and Meckler, J.F. (2013). Genome engineering at the dawn of the golden age. Annu. Rev. Genomics Hum. Genet. *14*, 135-158.

Segal, D.J., Dreier, B., Beerli, R.R., and Barbas, C.F., 3rd (1999). Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. Proc. Natl. Acad. Sci. USA *96*, 2758-2763.

Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Koksch, B., Lund, C.V., Magnenat, L., Valente, D., et al. (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry *42*, 2137-2148.

Segal, D.J., Goncalves, J., Eberhardy, S., Swan, C.H., Torbett, B.E., Li, X., and Barbas, C.F., 3rd (2004). Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. J. Biol. Chem. *279*, 14509-14519.

Segal, D.J., Crotty, J.W., Bhakta, M.S., Barbas, C.F., 3rd, and Horton, N.C. (2006). Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. J. Mol. Biol. *363*, 405-421.

Smith, M.C., and Thorpe, H.M. (2002). Diversity in the serine recombinases. Mol. Microbiol. *44*, 299-307.

Soldner, F., Laganiere, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318-331.

Sun, N., Liang, J., Abil, Z., and Zhao, H. (2012). Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. Mol. Biosyst. *8*, 1255-1263.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. *11*, 636-646.

Voigt, K., Gogol-Doring, A., Miskey, C., Chen, W., Cathomen, T., Izsvak, Z., and Ivics, Z. (2012). Retargeting sleeping beauty transposon insertions by engineered zinc finger DNA-binding domains. Mol. Ther. *20*, 1852-1862.

Wilber, A., Tschulena, U., Hargrove, P.W., Kim, Y.S., Persons, D.A., Barbas, C.F., 3rd, and Nienhuis, A.W. (2010). A zinc-finger transcriptional activator designed to interact with the gamma-globin gene promoters enhances fetal hemoglobin production in primary human adult erythroblasts. Blood *115*, 3033-3041.

Wu, H., Yang, W.P., and Barbas, C.F., 3rd (1995). Building zinc fingers by selection: toward a therapeutic application. Proc. Natl. Acad. Sci. USA *92*, 344-348.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordonez, A., Hannan, N.R., Rouhani, F.J., et al. (2011). Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391-394.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat. Biotechnol. *29*, 149-153.

Zou, J., Mali, P., Huang, X., Dowey, S.N., and Cheng, L. (2011). Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. Blood *118*, 4599-4608.