

De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks

Magdy M. Mahfouz, Lixin Li, Md. Shamimuzzaman, Anjar Wibowo, Xiaoyun Fang, and Jian-Kang Zhu¹

Center for Plant Stress Genomics and Technology, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

Contributed by Jian-Kang Zhu, December 26, 2010 (sent for review December 19, 2010)

Site-specific and rare cutting nucleases are valuable tools for genome engineering. The generation of double-strand DNA breaks (DSBs) promotes homologous recombination in eukaryotes and can facilitate gene targeting, additions, deletions, and inactivation. Zinc finger nucleases have been used to generate DSBs and subsequently, for genome editing but with low efficiency and reproducibility. The transcription activator-like family of type III effectors (TALEs) contains a central domain of tandem repeats that could be engineered to bind specific DNA targets. Here, we report the generation of a Hax3-based hybrid TALE nuclease with a user-selected DNA binding specificity. We show that the engineered TALE nuclease can bind to its target sequence in vitro and that the homodimeric TALE nuclease can cleave double-stranded DNA in vitro if the DNA binding sites have the proper spacing and orientation. Transient expression assays in tobacco leaves suggest that the hybrid nuclease creates DSB in its target sequence, which is subsequently repaired by nonhomologous end-joining repair. Taken together, our data show the feasibility of engineering TALE-based hybrid nucleases capable of generating site-specific DSBs and the great potential for site-specific genome modification in plants and eukaryotes in general.

molecular scissors | nontransgenic mutagenesis | artificial activators

Substantial improvements in the quality and yield of current crops are needed to meet the increasing food demand of the growing world population. These improvements will require the development of new approaches for agricultural biotechnology, which include targeted genome modification (1). The ability to perform targeted genome modification in plants would facilitate efficient and robust addition, deletion, activation, and inactivation of genes. A key requirement for the precise modification of the genome is the ability to introduce double-strand breaks (DSBs) at preselected and defined loci (2). Genomic DSBs are repaired by one of two widely conserved repair pathways, namely nonhomologous end joining (NHEJ) and homologous recombination (HR) (2). NHEJ repair joins the ends of broken strands and results in small deletions or insertions. HR, however, requires a homologous DNA segment as a template to copy the information across the break. The generation of targeted DSBs requires proteins that possess high fidelity of DNA recognition and provide site-specific cleavage (3, 4).

FokI protein, which is a type II restriction endonuclease produced by *Flavobacterium okeanokoites*, has an N-terminal DNA binding domain and a C-terminal nonspecific cleavage domain (5). The DNA binding domain recognizes the nonpalindromic 5' GGATG³ sequence, and the cleavage domain creates cleavages 9 and 13 bp downstream of the recognition sequence. FokI enzyme dimerizes in solution on binding to its DNA target sequence in the presence of divalent metal ions (6). The FokI nonspecific cleavage domain has been used in combination with other DNA binding domains (7).

Zinc finger nucleases (ZFNs) are hybrid proteins created by fusing a zinc finger DNA binding domain to the nonspecific cleavage domain of FokI endonuclease (5). Many studies have

shown that ZFNs can be engineered to target a specific and user-selected DNA sequence (8). The FokI cleavage domain must dimerize to cleave the DNA target. To achieve dimerization, two ZFNs must bind to the opposite DNA strands in a tail to tail orientation, with proper spacing (spacer length) between the two binding sites (9). Several modular assembly and selection strategies have been used to generate zinc fingers that bind a specific DNA sequence (10). Zinc finger modules can be assembled that recognize up to a 24-bp DNA with a 4- to 7-bp spacer and hence, that recognize a single locus in the genome. On the binding and dimerization of the two ZFNs to their target sequences, the nuclease cleavage domain creates a DSB in the spacer region (11). The requirement of the two binding events with the correct orientation and proper spacing permits specific targeting of the unique recognition sites. A site-specific DSB into any locus of interest can be subsequently repaired by the NHEJ pathway, which leads to occasional loss or gain of the genetic information and thereby, could be used to introduce small insertions or deletions (12). The outcome of the repair process also could disrupt the gene function.

ZFNs have been used to manipulate the genome of many organisms, including zebrafish (13), *Caenorhabditis elegans* (14), and plants (15, 16). Generating such specific ZFNs, however, is laborious and has had limited efficiency and success (17). Replacing the zinc finger domains with highly specific DNA binding domains that can be generated reproducibly would significantly improve the use of engineered nucleases as powerful tools for targeted genome modification.

Transcription activator-like effectors (TALEs) are used as virulence factors by phytopathogenic bacteria of the genus *Xanthomonas*; these bacteria use the type III secretion system to inject virulence factors into plant cells (18). After they are in the host cell, TALEs translocate to the nucleus, bind to their DNA targets, and mimic the eukaryotic transcription factors to reprogram host gene expression, leading to developmental changes and disease or resistance symptoms (19, 20). TALE proteins contain structural features and domains that include a nuclear localization signal (NLS), an N-terminal translocation signal, an acidic activation domain, and a central repeat domain that binds DNA. TALEs differ mainly in the number and order of repeats in the DNA binding domain (21). Each repeat is about 34–35 aa in length. Amino acids at positions 12 and 13 of each repeat constitute a repeat variable diresidue (RVD) that dictates the specificity of the repeat to recognize one nucleotide in the target sequence (21, 22). The one to one correspondence between the RVDs of the TALEs and the nucleotides in the target DNA

Author contributions: M.M.M. and J.-K.Z. designed research; M.M.M. and L.L. performed research; M.M.M., L.L., M.S., A.W., and X.F. contributed new reagents/analytic tools; and M.M.M. and J.-K.Z. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. E-mail: jian kang.zhu@kaust.edu.sa.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1019533108/-DCSupplemental.

sequence was elucidated by bioinformatics methods and experimental testing, and the code was deciphered and used to identify the targets of many TALEs (22). Moreover, the code was successfully tested using in vitro assembled TALEs with arbitrary RVDs and predicted DNA target sequences (21). According to this code, each RVD recognizes one or more nucleotide [asparagine isoleucine (NI) → A, histidine aspartic acid (HD) → C, asparagine glycine (NG) → T, asparagine asparagine (NN) → A, G and asparagine serine (NS) → A, C, G, and T] (21). It was recently reported that the NK RVD mediates specific interaction with the nucleotide G (23).

We reasoned that this cipher could be used to manipulate the RVDs of TALEs to generate TALEs with different DNA binding specificities for precise genome manipulations. Hax3 TALE is produced by the *Brassicaceae* pathogen *X. campestris* pv. *armoraciae* strain 5 (24). Hax3 is a member of the AvrBs3 family of TALEs. Hax3 shows the typical structural features of TALEs and has a DNA binding domain composed of 11.5 repeat units. The 11.5 RVDs of Hax3 recognize a 12-bp effector binding element (EBE) with sufficient binding strength to fully activate its target promoter (21).

We report here the use of Hax3 TALE as a scaffold to generate a DNA binding domain with a different DNA binding specificity. The engineered DNA binding domain was fused to a FokI cleavage domain to generate a de novo Hax3 TALE-based hybrid nuclease. We show the activity of the de novo Hax3 TALE-based hybrid nuclease in vitro and show its ability to create a DSB in vivo when transiently expressed in tobacco leaves. Our data suggest the feasibility of using engineered TALE-based hybrid nucleases as molecular scissors to create DSBs. Engineered TALE-based hybrid nucleases should be very useful for targeted genome modifications.

Results

Design and Construction of a de Novo Hax3 TALE-Based Hybrid Nuclease. The specificity of TALEs depends on the RVDs in the repeat region and their specific order (22). We selected the Hax3 effector as a scaffold to generate our hybrid TALE-based nuclease, because the protein is small relative to other TALE proteins and contains 11.5 repeats in the DNA binding region, which is a sufficient number of repeats to confer strong binding to and full activation of the promoter target (Fig. 1A) (21). However, the

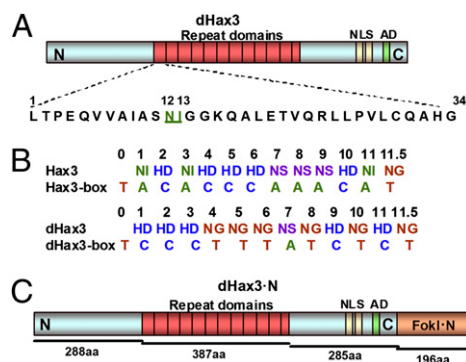


Fig. 1. Structural representation of Hax3 and dHax3.N proteins and their DNA target specificity. (A) Hax3 TALE contains a central repeat domain shown in red, a nuclear localization signal (NLS) shown in yellow, and a transcriptional activation domain (AD) shown in green. The central repeat domain consists of 11.5 repeat units, and each repeat unit is composed of 34 aa; a representative first repeat sequence is shown below, and the RVD is shown in blue. (B) RVDs of the 11.5 repeat units of Hax3 and dHax3 with their DNA target sequences shown below. (C) Structural representation of the dHax3.N showing the fusion of 196 aa of the FokI cleavage domain at the C terminus of dHax3.

sequence of the natural *Hax3* cDNA (GenBank accession no. AY993938.1) is GC-rich (65%), and the 11.5 repeats in the DNA binding region are nearly identical, which may complicate the PCR-based mutagenesis of these repeats (24). We, therefore, performed a codon optimization of the original *dHax3* cDNA for in planta expression. The codon-optimized *Hax3* sequence, herein referred to as *dHax3*, has a GC content of only 47%, and the 102-bp repeats share less homology than those of *Hax3* cDNA. The natural Hax3 TALE binds to the DNA box that contains the TACACCCAAACAT sequence (Fig. 1B) (21). The 11.5 repeats of dHax3 were modified, and therefore, it binds a 12-bp sequence (TCCCTTTATCTCT) in the *RD29A* promoter (Fig. 1B) (25). It should be noted that the target sequence is preceded by the nucleotide T at the repeat 0 (Fig. 1B). The presence of the T nucleotide at repeat 0 was shown to be essential for TALE activation of the respective promoters of their target genes, and it may be important for the binding of TALE to the DNA target (21).

To generate a hybrid dHax3 nuclease, we fused in frame the nucleotide sequence that corresponds to 196 aa of the FokI C-terminal cleavage domain to the full-length *dHax3* sequence (*SI Appendix*) (26). The hybrid dHax3 nuclease, herein referred to as dHax3.N, has 1,157 aa residues and an estimated molecular mass of 124 kDa (Fig. 1C). dHax3.N was designed so that the dHax3 provides the DNA binding specificity, whereas the cleavage domain (N) of the FokI protein provides the nuclease activity. The in silico-designed *dHax3.N* DNA sequence was synthesized in fragments that were fused together and cloned into a pUC19 vector and subsequently, into a pENTR/D gateway vector; the constructs were confirmed by sequencing. The *dHax3* and *dHax3.N* clones were subsequently subcloned in the pET32a expression vector to generate thioredoxin-6His fusion proteins (Trx.His.dHax3 and Trx.His.dHax3.N). The bacterially expressed Trx.6His.dHax3 and Trx.6His.dHax3.N proteins were shown to migrate in SDS protein gel electrophoresis according to their expected sizes (*SI Appendix* and Fig. S1).

dHax3 and dHax3.N Hybrid Nucleases Bind to Their DNA Target Sequence In Vitro and Activate Target Gene Expression In Vivo. We tested the ability of the bacterially expressed dHax3 and dHax3.N proteins to bind to the 12-bp EBE in vitro. The Trx.6his-tagged dHax3 and dHax3.N proteins were used in EMSA with biotin-labeled double-stranded oligonucleotides containing the EBE or its mutated versions (Fig. 2A). The EMSA showed that the Trx.6His.dHax3 and Trx.6His.dHax3.N proteins are capable of binding to the EBE target sequence. The binding of the dHax3 and dHax3.N proteins is shown to be competitively reduced by the addition of unlabeled double-stranded oligonucleotides of the same sequence (Fig. 2B and C). These data also indicate that the thioredoxin tag does not impede the binding of the dHax3 and dHax3.N proteins to the EBE target sequence (Fig. 2B and C).

To test whether dHax3 and dHax3.N proteins bind to their target sequence in vivo to activate gene transcription, we subcloned dHax3 and dHax3.N in a binary vector under the control of cauliflower mosaic virus 35S (35S) promoter. The 35S::dHax3 and 35S::dHax3.N binary clones were transformed into *Agrobacterium tumefaciens* for transient expression assays in *Nicotiana benthamiana* plants. A 300-bp *RD29A* promoter was cloned upstream of the *uidA* cDNA in the pK2GWFS7 binary vector (27). The 35S::dHax3 and 35S::dHax3.N containing *Agrobacterium* were each coinfiltrated in *N. benthamiana* leaves with *Agrobacterium* that contained *RD29A::uidA*. Leaf discs of the infiltrated regions were collected 24 h postinfection (hpi) and kept overnight in beta-glucuronidase (GUS) staining buffer. The data show that both dHax3 and dHax3.N proteins can bind to and activate the *RD29A* promoter (Fig. 2D).

dHax3.N Hybrid Nuclease Cleaves In Vitro the DNA Target Sequence with a 16-bp Spacer Between EBES. We tested the ability of the dHax3.N protein to cleave its double-stranded DNA target in

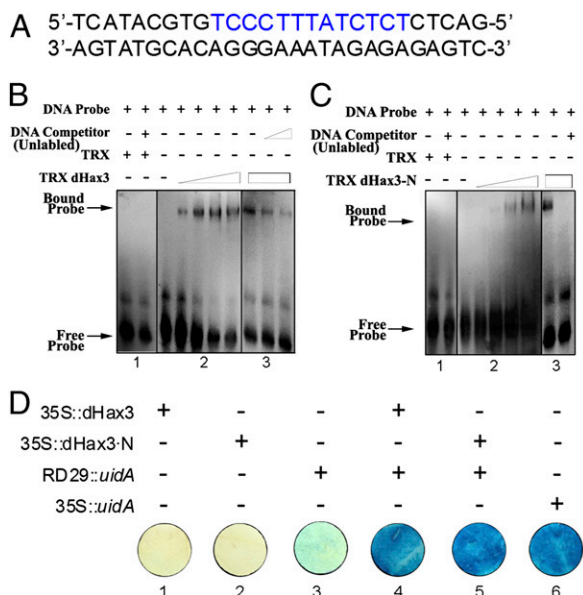


Fig. 2. Binding of dHax3 and dHax3.N to their target sequence and in vivo transcriptional activation. (A) Probe containing the EBE sequence (shown in blue) used in the EMSA. (B) EMSA showing dHax3 binding to its target sequence. Panel 1 shows that the thioredoxin tag does not bind to the biotin-labeled probe. Panel 2 shows that the dHax3 protein binds to the biotin-labeled probe sequence in a concentration-dependent manner. Panel 3 indicates that the unlabeled probe competitively reduces the binding of the biotin-labeled probe. (C) EMSA showing the binding of dHax3.N to its target DNA; panel 2 shows that dHax3.N binds to its target in a concentration-dependent manner, and panel 3 shows competition of binding by the unlabeled probe. (D) dHax3 and dHax3.N activate the *RD29A* promoter. Panels 1 and 2 serve as controls and indicate no *uidA* activity. Panel 3 indicates the background activity of *RD29A* promoter. Panels 4 and 5 show transcriptional activation of *RD29A* promoter by dHax3 and dHax3.N, respectively. Panel 6 shows a positive control of 35S::uidA. The experiment was repeated three times with similar results; 60 plants were used for each experiment (10 plants/panel).

vitro. Previous studies have shown that the nuclease activity of FokI enzyme and the zinc finger hybrid nucleases require dimerization for in vitro and in vivo activities (4, 11). To test the activity of our dHax3.N, we designed target sequences that contain two EBEs in a tail to tail orientation and that are separated by spacers with different lengths. We included a Small site in the spacer region between the two EBE target sequences to facilitate data analysis. pCRII/EBE6, pCRII/EBE10, and pCRII/EBE16 are three EBE clones with 6-, 10-, and 16-bp spacer regions, respectively. The bacterially expressed thioredoxin-tagged dHax3.N was used in the digestion reactions.

We used either NcoI or XcmI enzymes to linearize the pCRII/EBE clones. The linearized clones were gel-purified and used as substrates for the dHax3.N digestion reactions. The digestion reactions were performed with different buffers and a freshly purified dHax3.N enzyme for 30 min at 37 °C. No enzyme activity was observed using the EMSA DNA binding buffer (12 mM Tris-HCl, 60 mM KCl, 2.5% glycerol, 5 mM MgCl₂). Slight enzyme activity was observed using the FokI reaction buffer (20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9). The highest activity of the dHax3.N hybrid nuclease was found in a reaction buffer that contains 20 mM Tris-HCl (pH 7.4), 5 mM MgCl₂, 50 mM KCl, 5% glycerol, and 0.5 mg/mL BSA. The dHax3.N enzyme exhibits no activity when EBE with a 6-bp spacer-length clone (pCRII/EBE6) (Fig. 3C, EBE6 panel) was used and reduced activity when EBE with a 10-bp spacer-length clone was used (pCRII/

EBE10) (Fig. 3C, EBE10 panel). Our data show that the dHax3.N cleavage activity is the highest when the spacer length is 16 bp (pCRII/EBE16) (Fig. 3C, EBE16 panel).

We studied the effect of the EBE mutations in the pCRII/EBE16 clone on dHax3.N enzyme activity. We deleted the T nucleotide at the repeat 0 or mutated the T nucleotide to A at the repeat 12. We performed the digestion reactions in the optimized buffer as indicated above. Our data show that the deletion of T at repeat 0 significantly reduced the enzyme activity (Fig. 3C, EBE16ΔT0 panel). Interestingly, the replacement of T by A at repeat 12 also significantly reduced enzyme activity (Fig. 3C, EBE16T/A panel). The concentration of the cleavage products increased over time (Fig. 3D). We performed the cleavage reactions over a wide range of temperatures from 0 °C to 55 °C. Our data showed that the enzyme activity is highest between 23 °C and 37 °C. The enzyme showed a highly reduced activity at temperatures outside this range (Fig. 3E).

To determine the cleavage sites of dHax3.N, we gel-purified the 1,620- and 2,424-bp bands and used the M13 forward and M13 reverse primers for sequencing. The sequence data reveal that the enzyme cleaves within the spacer region (*SI Appendix* and Fig. S2).

dHax3.N Hybrid Nuclease Creates DSB in Its Target Sequence in *N. benthamiana* Transient Expression Assays.

To monitor the activity of dHax3.N in vivo in plant cells, we adopted a tobacco transient assay (28). First, we generated the EBE sequences in a tail to tail orientation separated by a 16-bp spacer that contains a TGA stop codon. The forward EBE sequence is preceded by an ATG start codon. The ATG.EBE.TGA.spacer.EBE is fused to the *uidA* reporter coding sequence (Fig. 4A). The ATG.EBE.TGA.spacer.EBE::uidA is generated by PCR amplification, and it is cloned in the pENTR/D vector and subsequently, subcloned in the pK2GW7 overexpression vector. When expressed in plant cells, the TGA stop codon would lead to a premature termination of *uidA* expression. If dHax3.N is expressed and active in vivo, however, it would bind to and generate a DSB in the ATG.EBE.TGA.spacer.EBE::uidA. The cellular repair machinery could subsequently repair the DSB, leading to the expression of the *uidA* reporter.

To test our hypothesis, we transformed our binary constructs 35S::ATG.EBE.TGA.spacer.EBE::uidA, 35S::uidA, 35S::dHax3, and 35S::dHax3.N into *A. tumefaciens* and then, coinfiltrated or infiltrated separately into *N. benthamiana* leaves. The infiltrated leaf discs were collected 48 hpi and used for GUS assays.

The 35S::ATG.EBE.TGA.spacer.EBE::uidA construct did not show activity, indicating that the TGA stop codon in the spacer region led to the premature termination of the reporter (Fig. 4B, panel 1). As a second control, 35S::ATG.EBE.TGA.spacer.EBE::uidA was coinfiltrated with the 35S::dHax3 construct, and no *uidA* activity was detected in the GUS assay of the leaf discs (Fig. 4B, panel 2).

However, when the same construct (35S::ATG.EBE.TGA.spacer.EBE::uidA) was coinfiltrated with 35S::dHax3.N, *uidA* activity was clearly shown, indicating the formation of DSBs in the spacer region and subsequent repair by the cellular repair machinery to remove the stop codon to allow the expression of the *uidA* reporter (Fig. 4B, panel 3). When the positive control 35S::uidA was infiltrated into leaves, *uidA* activity was evident (Fig. 4B, panel 4). These data indicate that a TALE-based hybrid nuclease can create DSBs in vivo when expressed transiently with their target sequence in *N. benthamiana* leaves.

Discussion

In this study, we showed the feasibility of the de novo engineering of a TALE hybrid nuclease with user-selected DNA binding specificity and in vitro and in vivo nuclease activity. We used the Hax3 TALE as a scaffold to build a hybrid nuclease and used its tandem repeat units as a DNA binding module to en-

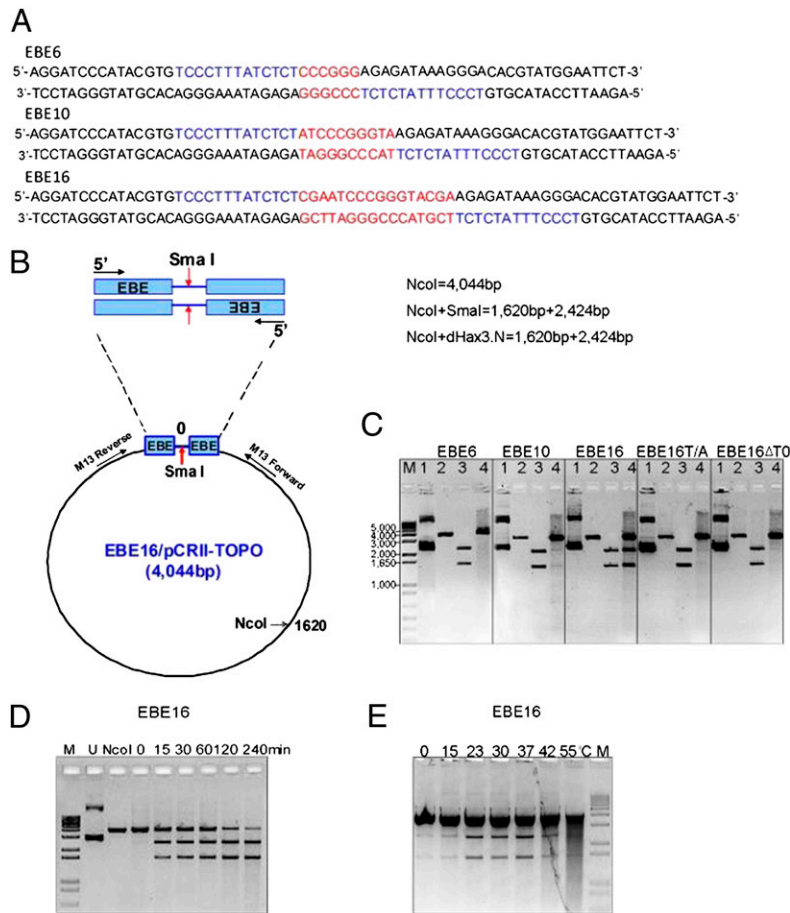


Fig. 3. In vitro digestion of DNA target with spacers of different lengths. (A) DNA target sequence with 6-, 10-, and 16-bp spacer EBE sequences are shown in blue, and the spacer sequences are shown in red. (B) Representation of the circular pCRII-TOPO plasmid containing two EBE elements in a tail to tail orientation and the predicted cleavage sites for Sma I and Nco I enzymes. (C) In vitro digestion of DNA targets with different spacer lengths. In all EBE panels, lane 1 is undigested plasmid. Lane 2 is Nco I-linearized plasmid. Lane 3 is digested with Nco I and Sma I. Lane 4 is 500 ng Nco I-linearized plasmid digested with 50 ng dHax3.N for 30 min at 37 °C. Lane 4 in panel EBE6 does not have the predicted digestion products, indicating no cleavage activity by dHax3.N with a 6-bp spacer between the two EBEs. Lane 4 in the EBE10 panel shows minimal activity of the enzyme, resulting in a very small amount of the predicted products. Lane 4 of panel EBE16 shows substantial activity of dHax3.N, indicating an optimal spacer length. In lane 4 of panel EBE16 T/A, the replacement of T by A in the last nucleotide in the target sequence results in minimal or no activity of dHax3.N. In lane 4 of EBE16ΔT0, the removal of T in the target sequence that corresponds to repeat 0 leads to minimal or no dHax3.N activity. M, 1-kb marker. (D) Kinetics of cleavage of EBE16 by 50 ng dHax3.N. U, undigested plasmid; Nco I, Nco I-linearized DNA. (E) Effect of temperature on dHax3.N activity.

gineer a desired DNA binding specificity. We selected the Hax3 TALE because of its moderate size, but it still has a sufficient number of tandem repeat units to bind and transcriptionally activate the target sequence (21). We showed that a de novo-engineered dHax3 TALE and its hybrid nuclease bind to the target sequence in vitro and in vivo and activate gene expression in tobacco transient assays. Moreover, fusion of the FokI nuclease domain to the de novo-engineered TALE does not disrupt the DNA binding ability of the tandem repeat units, which suggests great possibilities for various biotechnological applications (29). The tandem repeat units of TALEs could be used as DNA binding modules to engineer artificial transcriptional activators and repressors, DNA methylases and demethylases, recombinases, and nucleases that target chromosomal loci of interest.

The dimerization of the nuclease domain of native FokI and ZFN is required for the enzyme-cleavage activity. We designed a target site that contains two EBEs that have a tail to tail orientation and that are separated by a spacer of variable lengths. The purified dHax3.N protein showed the highest cleavage activity with a 16-bp spacer between the EBEs. Previous reports have shown that the linker length between the DNA binding domain and the nuclease domain determines the spacer length

between the EBE target sites (30). The native FokI nuclease has a 15-aa linker and cuts at 9 and 13 bp downstream of the target site on the sense and antisense strands, respectively. Also, a ZFN with an 18-aa linker length exhibits the highest cleavage activity when the target sites are separated by 8 bp (30, 31). ZFNs with shorter linkers cleave target sites with shorter spacers between the two EBEs (31). If required, it may be possible to generate TALE hybrid nucleases with the ability to cleave within a shorter spacer between the EBEs by reducing the linker length. In the present study, we used the full-length TALE, and the linker length between the DNA binding domain and the nuclease domain is 286 aa. The longer linker in our hybrid nuclease may explain why its activity was highest with a 16-bp spacer rather than a 10-bp spacer. It is important to determine the shortest part of TALE that could function as a DNA binding module and the required linker length to develop hybrid nucleases for applications in different organisms and cell types, because the size of the hybrid nuclease might be a critical factor.

To study the effect of the T nucleotide at repeat 0, we deleted this nucleotide from the target sequence to generate pCRII/EBE16ΔT and used this clone for in vitro digestion. The deletion of this nucleotide significantly reduces the cleavage activity of the

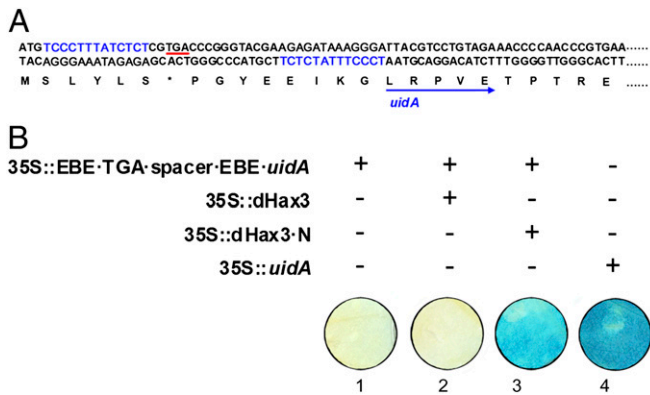


Fig. 4. dHax3.N activity in vivo. (A) The sequence of the EBE target is preceded by the initiation codon (ATG) and contains a TGA stop codon in the spacer region. The EBE target sequence is fused in frame to the *uidA* coding sequence. (B) dHax3.N activity in tobacco transient assays. Panels 1 and 2 show no activity of 35S::EBE.TGA.spacer.EBE.*uidA* or 35S::dHax3, respectively. Panel 3 shows *uidA* activity when 35S::EBE.TGA.spacer.EBE.*uidA* was coinfiltrated with 35S::dHax3.N. Panel 4 is a positive control that shows the activity of 35S::uidA. The experiment was repeated three times with similar results; 40 plants were used for each experiment (10 plants/panel).

dHax3.N enzyme, which is consistent with the importance of this nucleotide for the binding of the TALE to the DNA target (32).

Additionally, we showed the cleavage activity of dHax3.N enzyme in vivo. The EBE target sites preceded by the ATG start codon with a 16-bp TGA-containing spacer fused in frame with the *uidA* coding sequence was used as a cleavage target.

Our data suggest that the 16-bp spacer is sufficient for dHax3.N binding and subsequent cleavage of the target site in vivo. dHax3.N creates DSBs within the predicted target sequence in the spacer region as shown by the sequencing trace data of the cleavage products (SI Appendix). Our results suggest that the DSBs are subsequently repaired by the NHEJ repair machinery, leading to elimination of the TGA stop codon and resulting in the production of functional copies of *uidA* reporter. These data show that the dHax3.N hybrid nuclease has in vivo activity and reveal a great potential of using these hybrid nucleases to generate site-specific chromosomal breaks. The data indicate that the de novo TALE-based hybrid nucleases could be used to introduce site-specific heritable mutations in the genome, thereby obviating the need to stably transform plants to generate such mutations. Moreover, this technology could be developed for gene-targeting purposes. The rate-limiting step for gene targeting is the generation of site-specific DSBs. Although homing endonuclease *I-SceI* and ZFNs have been used to create genomic DSBs in a variety of organisms, including plants (33), they lack the required flexibility, adaptability, and reproducibility to be used as tools for routine genome engineering.

The data reported here and recently published by Christian et al. (34) and Li et al. (35) indicate that the TALE cipher and the use of the tandem repeat units as DNA binding modules offer the required flexibility and adaptability. Christian et al. (34) have shown that AvrBs3 and PthXo1 TALEs lacking the activation domain and fused to the FokI nuclease domain could function as homodimers and heterodimers to achieve DNA cleavage in a yeast assay (34). Li et al. (35) showed that the fusion of the FokI cleavage domain to the C or N terminus of the AvrXa7 and PthXo1 TALEs produces a functional chimeric nuclease with the ability to cleave the DNA target in vitro and in yeast assay (35).

We predict that TALE-based hybrid nucleases will meet the criteria for creating DSBs in that they can be specific, flexible, and adaptable. The DNA binding specificity and flexibility of the TALE proteins coupled with the robust cleavage of the FokI

catalytic domain make these hybrid nucleases ideal tools for generating DSBs for genome engineering. These de novo-engineered hybrid nucleases could be used to achieve gene disruptions, corrections, and additions (Fig. 5).

It remains to be shown whether a de novo-engineered TALE nuclease could be used to create DSBs in a chromatin context. Our transient expression data in tobacco leaves indicate that dHax3.N has cleavage activity in vivo in plants. Next, it must be determined whether a TALE hybrid nuclease designed to bind to a specific chromosomal locus in either homodimer or heterodimer fashion would be active. The ability of these hybrid nucleases to create chromosomal DSBs in a chromatin context in vivo will undoubtedly usher in a new era for genome engineering in plants and eukaryotes in general.

Materials and Methods

dHax3 TALE Sequence Optimization and Synthesis and Vector Construction. The dHax3 cDNA was optimized to reduce the homology between the 102-bp repeat units, and the codons were optimized for in planta expression. The *dHax3* and *dHax3.N* cDNAs (SI Appendix) were synthesized by the BlueHeron Bio in a pUC19 vector. The cDNAs were amplified and cloned in the pENTR/D vector to generate pENTR/*dHax3* and pENTR/*dHax3.N* gateway entry clones. These clones were confirmed by sequencing and found to be identical to the sequence in the text (SI Appendix). LR recombination reactions were performed between the entry clones and gateway-compatible pET32a expression vector, as described by the manufacturer's instructions, to generate pET32a.dHax3 and pET32a.dHax3.N, respectively. The expression clones were transformed into *Escherichia coli* BL21, and the protein expression was induced at 25 °C for 5 h with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The His.dHax3, His.dHax3.N, TRX.dHax3, and TRX.dHax3.N proteins were purified using Qiagen Ni-NTA agarose resin according to the manufacturer's instructions. LR reactions were performed between the pENTR/dHax3 and pENTR/dHax3.N entry clones and the pK2GW7 gateway compatible binary vector to generate overexpression clones pK2GW7/dHax3 and pK2GW7/dHax3.N. These expression clones were transformed into *A. tumefaciens* GV3101 and used for transient expression in tobacco leaves.

EMSA. For dHax3 and dHax3.N DNA binding studies, TRX.dHax3 and TRX.dHax3.N were purified as described above, and the protein concentration was measured with a Bradford protein assay kit (BioRad). Complementary pairs of 5' biotin-labeled or nonlabeled oligonucleotides were annealed and used as probes for the binding studies. All of the EMSA reactions were performed using the Lightshift Chemiluminescent EMSA kit (Pierce) according to the manufacturer's instructions. EMSA binding-reaction buffer contained 12 mM Tris-HCl, 60 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM MgCl₂, 50 ng/μL poly(dI.dC), 0.05% Nonidet P-40, 0.2 mM EDTA, 35 fmol biotin-labeled DNA, 10 pmol unlabeled DNA, and 30–480 fmol TRX.His fusion proteins. The EMSA binding reactions were kept on ice for 10 min before the biotin-labeled probe was added. Gel electrophoresis was performed using 8% Tris/Borate/EDTA (TBE) native ready-made gels from Invitrogen. Blotting was performed on a positively charged nylon membrane, and the membrane was cross-linked using CL-1000 UV cross-linker for 30 s.

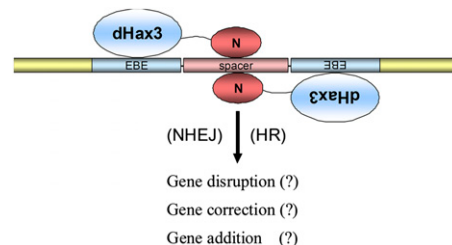


Fig. 5. Diagram of the dHax3.N hybrid nuclease with its DNA target and potential applications. dHax3 TALE provides a DNA binding module, and the FokI cleavage domain provides the nuclease activity. Two monomeric subunits of dHax3.N are required for the cleavage activity. DSBs created by dHax3.N could be used to achieve gene disruption, correction, and addition mediated by NHEJ or HR pathways.

dHax.3N in Vitro DNA Cleavage Reactions. dHax.3N EBEs were cloned in a tail to tail orientation in sense and antisense strands, respectively. The cloning of EBEs was achieved by the synthesis of the EBEs complementary ssDNA (primer sequences in S3) with 3'A overhangs to facilitate TA cloning. The EBE ssDNA was annealed and cloned into the TOPO TA cloning vector (Invitrogen), and the EBE clones were identified by sequencing.

The pCRII/EBE6, pCRII/EBE10, and pCRII/EBE16 clones were digested to completion by NcoI, and 500 ng linearized EBE clones were used in the digestion reactions with Trx.His.dHax.3N-purified protein. The digestion reactions were performed in a buffer that contained 20 mM Tris-HCl (pH 7.4), 5 mM MgCl₂, 50 mM KCl, 5% glycerol, and 0.5 mg/mL BSA. The reactions were incubated for 30 min at 37 °C and then analyzed by DNA gel electrophoresis.

To study the effect of mutations in the DNA target on the binding and cleavage activity of the dHax.3N enzyme, we generated two mutant clones by removing the T nucleotide that corresponds to repeat 0 and replacing the last T that corresponds to repeat 11.5 with A. Primers used to produce these clones are shown in [SI Appendix](#).

- Pennisi E (2010) Sowing the seeds for the ideal crop. *Science* 327:802–803.
- Weinthal D, Tovkach A, Zeevi V, Tzfira T (2010) Genome editing in plant cells by zinc finger nucleases. *Trends Plant Sci* 15:308–321.
- Cathomen T, Joung JK (2008) Zinc-finger nucleases: The next generation emerges. *Mol Ther* 16:1200–1207.
- Carroll D (2008) Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Ther* 15:1463–1468.
- Kim YG, Shi Y, Berg JM, Chandrasegaran S (1997) Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene* 203:43–49.
- Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* 95:10570–10575.
- Kim YG, Smith J, Durgesha M, Chandrasegaran S (1998) Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol Chem* 379:489–495.
- Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. *Nat Biotechnol* 23:967–973.
- Szcepek M, et al. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* 25:786–793.
- Wright DA, et al. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* 1:1637–1652.
- Vanamee ES, Santagata S, Aggarwal AK (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol* 309:69–78.
- Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11:636–646.
- Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 26:695–701.
- Carroll D, Beumer KJ, Morton JJ, Bozas A, Trautman JK (2008) Gene targeting in *Drosophila* and *Caenorhabditis elegans* with zinc-finger nucleases. *Methods Mol Biol* 435:63–77.
- Wright DA, et al. (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* 44:693–705.
- Townsend JA, et al. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459:442–445.
- Ramirez CL, et al. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* 5:374–375.
- Boch J, Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu Rev Phytopathol* 48:419–436.
- Gürlebeck D, Szurek B, Bonas U (2005) Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. *Plant J* 42:175–187.
- Römer P, Recht S, Lahaye T (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc Natl Acad Sci USA* 106:20526–20531.
- Boch J, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509–1512.
- Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501.
- Morbiter R, Römer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci USA* 107:21617–21622.
- Kay S, Boch J, Bonas U (2005) Characterization of AvrBs3-like effectors from a Brassicaceae pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol Plant Microbe Interact* 18:838–848.
- Yamaguchi-Shinozaki K, Shinozaki K (1993) Characterization of the expression of a desiccation-responsive rd29 gene of *Arabidopsis thaliana* and analysis of its promoter in transgenic plants. *Mol Gen Genet* 236:331–340.
- Carroll D, Morton JJ, Beumer KJ, Segal DJ (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nat Protoc* 1:1329–1341.
- Karimi M, Depicker A, Hilson P (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol* 145:1144–1154.
- Zhang F, et al. (2010) High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc Natl Acad Sci USA* 107:12028–12033.
- Voytas DF, Joung JK (2009) Plant science. DNA binding made easy. *Science* 326:1491–1492.
- Bibikova M, et al. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 21:289–297.
- Händel EM, Alwin S, Cathomen T (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: The inter-domain linker as a major determinant of target site selectivity. *Mol Ther* 17:104–111.
- Römer P, et al. (2010) Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol* 187:1048–1057.
- Maeder ML, et al. (2008) Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31:294–301.
- Christian M, et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186:757–761.
- Li T, et al. (2011) TAL nucleases (TALENs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* 39:359–372.

A. tumefaciens-Mediated Transient Expression and dHax.3N Hybrid Nuclease Activity Assays. To assay the activity of dHax.3N in planta, we adapted a transient assay in *N. benthamiana* leaves. The dHax.3N target was designed to include the fusion of EBE-stop codon.spacer-EBE fused in frame to *uidA* cDNA and cloned in a binary vector to generate pK2GW7(35S::EBE-TGA.spacer-EBE-GUS). The EBE-tga.spacer-EBE-GUS clone was generated by amplifying the *uidA* cDNA using the EBE EBE-tga-EBE-GUS-F and GUS-R primers ([SI Appendix](#)). The PCR product was cloned in the pENTR/D vector and subsequently cloned by LR reaction in the pK2GW7 overexpression vector. The pK2GW7/EBE-tga.spacer-EBE-*uidA* and pK2GW7/dHax.3N were transformed into *A. tumefaciens* GV3101 and coinfiltrated or infiltrated separately into tobacco leaves, and the infiltrated discs were collected 48 hpi. The GUS qualitative assay was performed by immersing leaf discs in GUS-staining solution (10 mM sodium phosphate, pH 7, 10 mM EDTA, 0.1% Triton X-100, 0.1% 5-bromo-4-chloro-3-indolyl- β -D-glucuronide, 1 mM potassium ferricyanide, 1 mM potassium ferrocyanide) at 37 °C for 24 h. The discs were cleared in ethanol.

ACKNOWLEDGMENTS. We thank John Klingler for helpful comments on the manuscript.