TAL Effectors: Customizable Proteins for DNA Targeting

Adam J. Bogdanove[1]* and Daniel F. Voytas[2,3]

Generating and applying new knowledge from the wealth of available genomic information is hindered, in part, by the difficulty of altering nucleotide sequences and expression of genes in living cells in a targeted fashion. Progress has been made in engineering DNA binding domains to direct proteins to particular sequences for mutagenesis or manipulation of transcription; however, achieving the requisite specificities has been challenging. Transcription activator–like (TAL) effectors of plant pathogenic bacteria contain a modular DNA binding domain that appears to overcome this challenge. Comprising tandem, polymorphic amino acid repeats that individually specify contiguous nucleotides in DNA, this domain is being deployed in DNA targeting for applications ranging from understanding gene function in model organisms to improving traits in crop plants to treating genetic disorders in people.

Cells use suites of DNA binding proteins to control the expression, replication, and transmission of the genetic material. These proteins typically recognize specific DNA sequences through DNA binding domains that tether the proteins to genomic sites where their activity is required, for example, to activate or repress gene transcription or to initiate DNA replication. Zinc fingers, helix-turn-helix motifs, and leucine zippers are some of the more prevalent protein folds, common across multiple kingdoms of life, that enable proteins to find specific genomic targets.

For discovery and application, biologists have sought to manipulate genetic information in cells by DNA targeting–engineering DNA binding domains with new sequence specificities and fusing them to proteins that modify DNA or its expression. The zinc finger domain, which predominantly recognizes nucleotide triplets, has been particularly widely used. Arrays of zinc fingers assembled to recognize targets of various lengths upstream of genes have been fused to transcriptional activation or repressor proteins to create artificial gene regulators (1), and sequence-specific zinc finger nucleases (ZFNs) have been created that, through directed chromosome cleavage, enable targeted mutagenesis and genome editing (2). Although considerable progress has been made in DNA targeting with

zinc fingers (3–5), their widespread adoption has been hindered by the resource intensive and empirical nature of achieving new DNA sequence specificities. This is due to a lack of known fingers for some nucleotide triplets and context

effects on the specificities of individual fingers in an array.

Transcription activator–like (TAL) effectors recognize DNA in an apparently modular fashion, described by us and others (6, 7), that is more amenable to DNA targeting: Tandem, polymorphic amino acid repeats in these proteins independently specify single, contiguous nucleotides in the DNA target (Fig. 1). Found as yet only in plant pathogenic bacteria, particularly members of the genus *Xanthomonas*, TAL ef-

fectors are, in fact, trans-kingdom, positive-acting transcription factors. They are injected into plant cells via the bacterial type III secretion system, imported into the plant cell nucleus, and targeted to effector-specific gene promoters (8, 9). TAL effector binding activates expression of downstream genes, which may contribute to bacterial colonization, symptom development, or pathogen dissemination [reviewed in (10, 11); see also Box 1]. Increasingly, scientists are exploiting the modularity of TAL effector–DNA recognition for DNA targeting to achieve control over the genetic material in vivo.

How Do TAL Effectors Recognize DNA?

The amino acid repeats of TAL effectors, located centrally in what we refer to as the targeting domain of the protein, are typically composed of 34 amino acids. However, variants with 33 or 35 amino acids are not uncommon, and the last repeat in the domain is truncated at 20 amino acids. Most TAL effectors have between 13 and 28 repeats (11). Polymorphism among the repeats is almost exclusively localized to a pair of residues at positions 12 and 13, called the repeat-variable di-residue (RVD). Different RVDs associate pref-
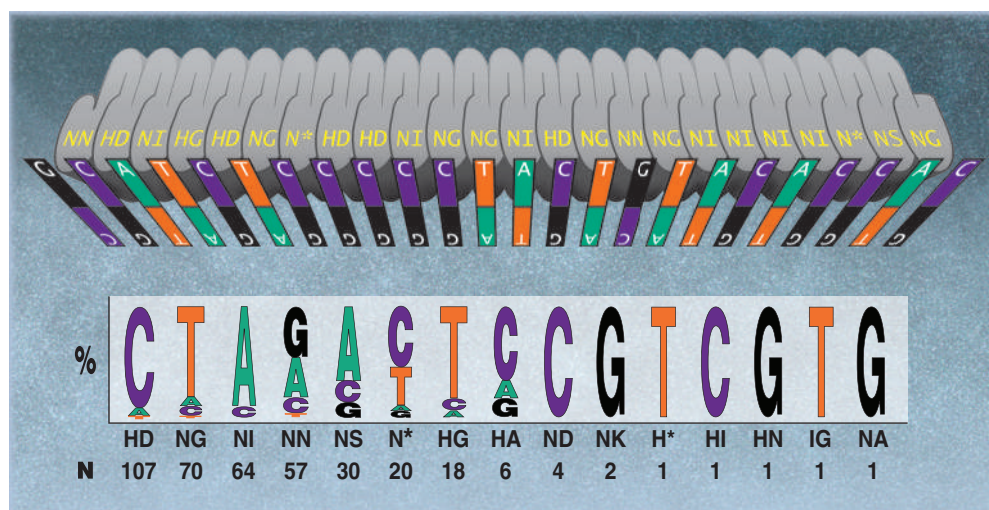
Fig. 1. TAL effector DNA recognition. (**Top**) DNA-targeting domain of TAL effector PthXo1 of *X. oryzae* and its target in the rice genome. TAL effector targeting domains contain a variable number of tandem, full-length repeats, typically 34 amino acids each, and a final truncated repeat of 20 amino acids. Each repeat, and the truncated repeat, displays a different pair of residues at positions 12 and 13 (the RVD; yellow text) that associates preferentially with one or more of the four nucleotides. (**Bottom**) Frequencies of RVD-nucleotide associations across 20 TAL effectors and their naturally occurring targets constitute a code that allows prediction and design. [Adapted from (7)]. An asterisk indicates that the residue at position 13 is missing, resulting in a 33–amino acid repeat.

erentially with different nucleotides, with the four most common RVDs (HD, NG, NI, and NN) (12) accounting for each of the four nucleotides (C, T, A, and G, respectively). Thus, the number of repeats (including the final, truncated repeat) and the string of RVDs they contain determine the length and nucleotide composition of those target sequences that are recognized. No structure has yet been reported for a TAL effector bound to DNA, but presumably RVDs make specific contacts with nucleotides (or base pairs) for target

[1]Department of Plant Pathology, 351 Bessey Hall, Iowa State University, Ames, IA 50011, USA. [2]Department of Genetics, Cell Biology and Development and Center for Genome Engineering, 321 Church Street Southeast, University of Minnesota, Minneapolis, MN 55455, USA. [3]Cellectis Plant Sciences, 1000 Westgate Drive, Suite 136, St. Paul, MN 55114, USA.

*To whom correspondence should be addressed. E-mail: ajbog@iastate.edu
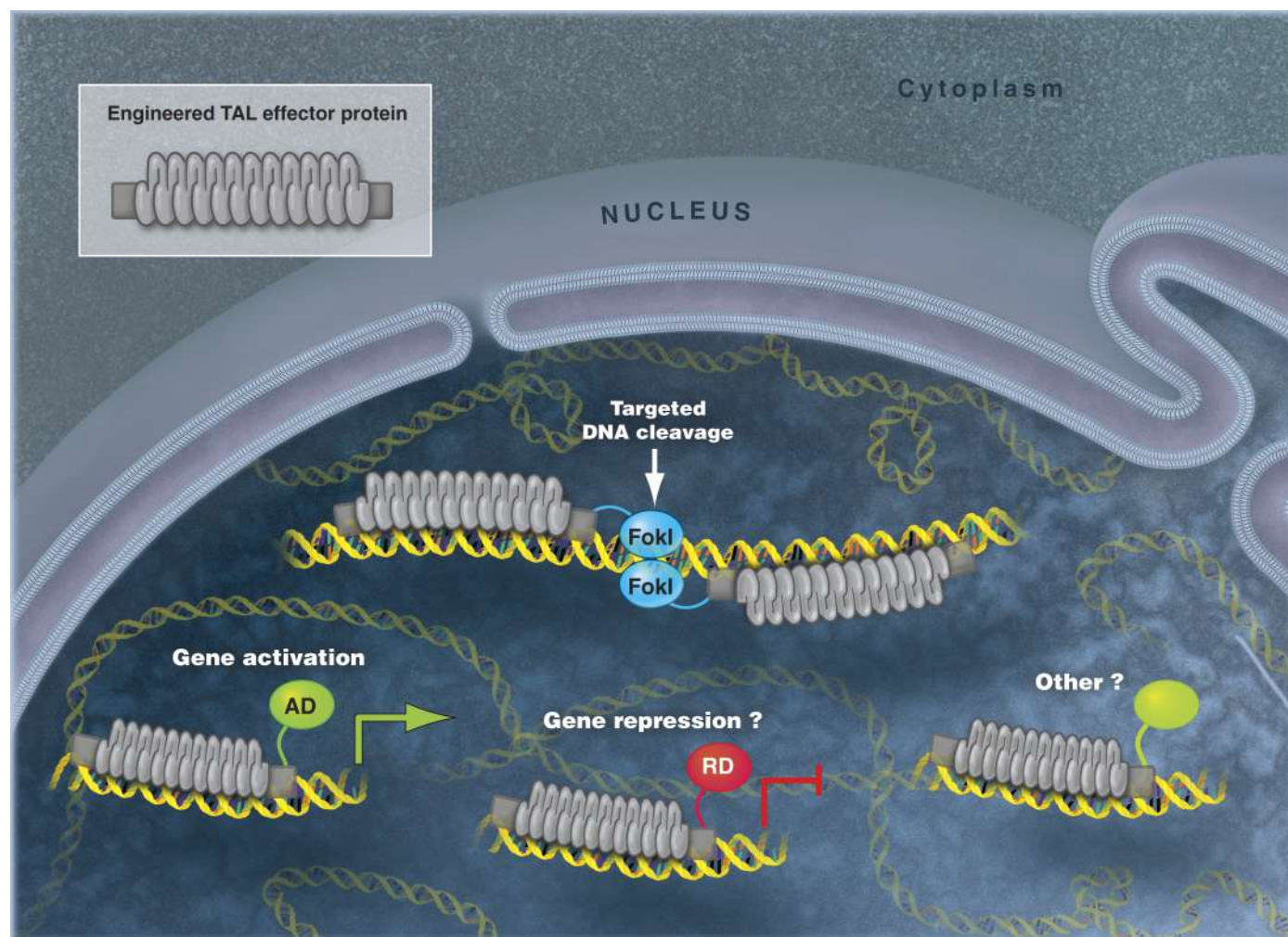
Fig. 2. Genomic control enabled by engineered TAL effector proteins. Fusion of TAL effector proteins to *Fok*I creates sequence-specific nucleases that enable targeted DNA cleavage for gene knockouts and genome editing. TAL effector proteins fused to transcriptional activation domains (AD) and putatively to repression domains (RD) provide artificial switches for gene regulation in vivo. TAL effector–based sequence-specific mutagens or chromatin-modifying proteins created by fusing TAL effectors to domains such as cytidine deaminases, histone acetyltransferases or deacetylases, or DNA methyltransferases can also be envisioned.

recognition. RVD-nucleotide associations are not exclusive, and most TAL effector/DNA pairs in nature contain mismatches; however, the most frequent associations constitute a code by which TAL effector binding sites can be predicted (*6*, *7*, *13–15*), target sites synthesized to bind particular TAL effectors (*6*, *14*), and custom TAL effector repeat arrays assembled to target DNA sequences of choice (*16–22*). In addition to the nucleotide sequence specified by the RVDs, naturally occurring DNA targets are uniformly preceded by a T, which is required for TAL effector activity (*6*, *7*). It has been proposed that the portion of the protein immediately preceding the repeat region interacts with and specifies this T because it shows predicted secondary structural similarity to a repeat, though it does not contain a recognizable RVD (*10*).

How Can TAL Effectors Be Used for Genome Engineering?

We and others showed that the TAL effector targeting domain can be used to direct the catalytic domain of the *Fok*I nuclease, as a fusion protein, to create site-specific DNA double-strand breaks (DSBs) (*17*, *18*, *20*, *23*). Because *Fok*I functions as a dimer, such TAL effector nucleases (TALENs) are designed in pairs that bind opposing DNA target sites separated by a spacer (Fig. 2), allowing the *Fok*I monomers to come together to create a DSB. The DSB, in turn, activates the cell's DNA repair pathways, which can be harnessed to create specific DNA sequence modifications at or near the break site (*2*). In nearly all cells, DSBs are repaired by one of two highly conserved processes. In nonhomologous end joining (NHEJ), the broken chromosome may be rejoined imprecisely, resulting in small insertions or deletions at the break site that can disrupt gene function. In homologous recombination (HR), the DNA surrounding the break site is replaced with a repair template of similar sequence. The sequence of the repair template can be modified or amended to swap in specific mutations or additional sequences (referred to as DNA editing). Genomic modifications based on both NHEJ and HR have been obtained with high frequency in a variety of plant and animal species using ZFNs and engineered homing endonucleases, the latter of which are mobile element-derived enzymes that recognize 12 to 40 base pairs (bp) and have also been engineered for new DNA sequence specificities (*24*).

We demonstrated TALEN activity on plasmid-borne targets in yeast using an assay in which the DNA recognition sequence is placed between two overlapping fragments of a reporter gene (*17*). If the target is cleaved, subsequent repair by single-strand annealing joins the overlapping fragments and reconstitutes the reporter, providing a quantitative readout of TALEN activity. Though TAL effectors with randomly assembled repeat arrays had been shown to function as transcriptional activators with corresponding synthesized target sequences (*6*), the yeast experiments demonstrated that repeat arrays can be customized to target specific sequences of interest, in this case gene sequences from *Arabidopsis thaliana* and zebrafish (*17*).

Another milestone was achieved when custom TALENs were shown to mediate both site-directed mutagenesis by NHEJ and gene targeting by HR at endogenous chromosomal loci in cultured human embryonic kidney cells (*18*). Efficiencies of NHEJ mutagenesis were as high as 25%, on par with those observed for ZFNs. In the 6 months since this milestone was achieved, TALENs have been used successfully in several experimental systems. TALEN-mediated gene targeting was demonstrated at five loci in human embryonic stem cells and induced pluripotent stem cells, again as efficiently as ZFNs (*25*). With regard to plants, TALENs were shown to cleave a transiently introduced, episomal target in leaves of tobacco (*20*), and NHEJ-mediated site-directed mutagenesis was achieved at an endogenous locus in *A. thaliana* protoplasts (*16*). In yeast, TALENs enabled high-efficiency gene replacement of several chromosomal loci (*26*). In *Caenorhabditis elegans* and its relative *C. briggsae*, for which gene targeting approaches had been lacking, TALEN or ZFN mRNA injected into gonads resulted in 3 to 5% of progeny with mutations in the endogenous target genes (*27*). Somatic cell and heritable gene knockouts in zebrafish have been made using TALENs (*28, 29*). And *IgM* knockout rats were generated by embryo microinjection of TALEN DNA or mRNA constructs (*30*).

A general concern in the use of customized targeting proteins for genome engineering is the possibility for mutagenesis at unintended sites. This is particularly the case in human therapeutics, in which off-target activity can have disastrous consequences (*31*). Even in plants and animals, which can be bred to remove unwanted mutations, the time and economic costs of doing so make specificity of paramount importance. TALEN specificity is affected by the spacer lengths between the two binding sites that permit cleavage. Depending on the TALEN architecture, functional spacer lengths from 6 to 40 bp have been reported, but for all of these architectures, the TALENs cleaved across a range of spacer lengths extending from 10 to 30 bp around the optima (*17, 18, 23, 32*). ZFNs, in contrast, have much more defined spacer length requirements (*33*). The range of functional spacer lengths for TALENs suggests that flexibility is provided by TAL effector protein sequences downstream of the targeting domain that allows the *Fok*I monomers to dimerize across different distances. Although several TAL effector derivatives with different N- and C- terminal truncations surrounding the targeting domain have been tested, the minimal portion of the larger protein that is essential for DNA binding has yet to be systematically delimited. An architecture comprising the minimal DNA binding domain and a length-optimized linker might result in a reduced spacer-length range that improves specificity.

The composition and number of repeats used may also affect specificity, because RVD-nucleotide preferences differ in their stringency and because the relative affinities of different RVDs for their preferred nucleotides, which are not yet known, might vary. However, the relative length of TALEN binding sites (13 nucleotides for each monomer, including the T that precedes the target sequence) is the minimum so far shown to work for TALENs, and the fact that TALENs function as pairs should make them highly specific in general. In the simplified theoretical example in which RVDs are assumed to have absolute specificity, given paired arrays of 15 RVDs separated by a spacer that might vary as much as 20 bp in length, even allowing up to three mismatches anywhere in each array, a TALEN would bind only once in 6.2 billion bp, twice the size of the human genome. Indeed, experimental evidence of TALEN specificity is beginning to accrue. In the study targeting endogenous loci in yeast, whole-genome sequencing of three strains treated with TALENs (and one with ZFNs) found no evidence of off-target mutagenesis (*26*). The study in human stem cells used SELEX (systematic evolution of ligands by exponential enrichment) to determine the diversity of nucleotide sequences bound by a given TALEN (*25*). Of 19 maximal-likelihood off-target sites surveyed in cells that had undergone TALEN-mediated gene editing, 17 remained wild-type, and the other two were disrupted 169-fold and 1140-fold less frequently than the intended target. In a study comparing TALEN and ZFN targeting of the human *CCR5* gene, TALENs were significantly less toxic to the treated cells than a ZFN, suggesting higher target specificity, and they discriminated the highly similar *CCR2* locus

approximately 10-fold better (*32*). In the heritable gene knockout study in zebrafish, rates of mutation at nine predicted off-target sites were no higher than rates in untreated controls (*29*). Finally, in the rat study, only one of nine sites showing sequence similarity to the target showed evidence of mutation (*30*). Thus, although it may ultimately prove difficult to specifically target some sequences—for example, those distinguished only by RVDs of low stringency or low affinity—there are grounds for optimism that off-target cleavage will not be a major barrier to the use of TALENs in genome engineering.

In What Ways Can TAL Effectors Be Used to Manipulate Gene Expression?

Several groups have successfully customized TAL effectors for specific gene activation, using either the native activation domain or, in its place, the VP16 activation domain of herpes simplex virus or its tetrameric derivative VP64 (*18, 19, 22, 34*) (Fig. 2). These designer TAL effectors (dTALEs) (*22*) function both in plants and human cells and commonly increase target gene expression by 20-fold or more. Perhaps not surprisingly, in plants, dTALEs with the native activation domain were more effective than those with VP16, and in human cells, the reverse was true (*34*). Targeted gene repression can also be envisioned in which a custom TAL effector targeting domain is fused to a transcriptional repressor domain (Fig. 2).

Because dTALEs function as monomers, their target specificity can be expected to be lower than that of TALENs. In fact, the first dTALE

Box 1: The TAL Effector DNA Recognition Code for Engineering Plant Resistance to *Xanthomonas*

Not all, but many *Xanthomonas* strains deploy TAL effectors during infection [reviewed in (*11*)]. TAL effector targets that play important roles in disease are called susceptibility (*S*) genes. Plants have evolved a handful of mechanisms to defend against TAL effector–wielding pathogens. Alleles of major *S* genes exist that confer resistance by means of a polymorphism in their promoter that prevents binding and activation by the corresponding TAL effector. Also, a polymorphism in a component of the transcription pre-initiation complex (PIC) was found in rice that confers resistance, presumably by preventing efficient recruitment of the PIC by TAL effectors for *S* gene activation. At least one member of the NB-LRR (nucleotide-binding–leucine-rich repeat) family of plant disease resistance (*R*) proteins recognizes a TAL effector, triggering defense. Finally, two genes, called executor *R* genes, have been identified whose transcriptional activation by a TAL effector blocks disease progression. The problem remains, though, that most of these mechanisms target just one TAL effector, so they may be effective only against particular strains. When incorporated into commercial plant varieties, these types of resistance can fail if strains arise that have alternative TAL effectors for *S* gene activation, target alternative *S* genes, or lack TAL effectors that might activate an executor *R* gene.

The TAL effector DNA recognition code, which has expedited the identification of *S* genes (*14, 15*) and executor *R* genes (*13*), opens up prospects for engineering such genes for durable and broad-spectrum resistance. Specifically, site-directed mutagenesis of the TAL effector binding sites in the promoters of a selection of *S* genes can be envisioned that would disarm a collective arsenal of TAL effectors present in a pathogen population. Further, executor *R* gene promoters could be engineered to trap not one, but multiple TAL effectors, chosen from among those most broadly conserved and important for virulence (*37*). The feasibility of the latter approach has been demonstrated in part in an episomal reporter gene assay in the model plant *Nicotiana benthamiana* (*13*). Most likely, it will be possible to make the required chromosomal modifications in crop plants with the use of TAL effector nucleases.

---

study in *Arabidopsis* predicted four total off-target sites for two dTALEs, each with two mismatches. The sequence space considered for off-target sites was not the entire genome, just the annotated promoter sequences. Only one predicted off-target site was activated by the corresponding dTALE. This is consistent with previous observations that different mismatches, or mismatches in different positions, can have distinct consequences; even single mismatches sometimes disrupt activity entirely (*10*). Also, the location of the binding site within the promoter and its chromatin state probably influences activity. Considering these constraints on functional binding, targeting specificity for dTALEs may not be as difficult to achieve in practice as it might appear, though a better understanding of mismatch tolerance, positioning requirements, and chromatin effects will be required.

Are There Constraints to Designing and Assembling Custom TAL Effector Constructs?

Whereas zinc finger protein engineering is constrained by the fact that neighboring fingers in an array influence specificity in an undefined manner, TAL effector customization appears to be free from this limitation. The successful targeting of large numbers of diverse DNA sequences using TALENs and dTALEs provides strong evidence of context-independent association of individual RVDs with individual nucleotides. However, a systematic study to examine this question and to determine other possible design constraints has yet to be reported. Among naturally occurring TAL effectors and corresponding target sites, we found no neighbor effects, but we did find positional and overall biases in nucleotide and RVD composition, which suggested requirements for design. In a diverse collection of genes from different organisms, we identified paired sequences that conformed to these biases, appropriately spaced for cleavage by a TALEN and each preceded by a T, on average, every 35 bp (*16*). The estimated frequency of targetable sites with the widely used context-dependent assembly platform for ZFNs is every 500 bp (*4*). Each of the 15 sites targeted with custom TALENs that conform to the naturally occurring biases were cleaved efficiently in yeast (*16*). Whether the biases indeed reflect structural requirements or are instead simply a relic of the shared phylogeny of these proteins remains to be determined. Functionality of several array and target sequence pairs that we and others tested that do not conform (*16*, *18*) argues in favor of the latter.

Designing for target specificity may, in some cases, be hampered by the lax specificity of RVD NN, which associates with A almost as frequently as with its preferred nucleotide, G (*7*). On the other hand, NN, and the less common NS, which appears to lack nucleotide specificity altogether (*7*), may provide a measure of flexibility in design to target a set of similar but not identical sequences (*22*). Two studies provided evidence that the rare RVD NK has better specificity for G

than NN does (*18*, *22*), but whether arrays with NK in place of NN bind with comparable affinity needs to be tested.

Design constraints can be explored with the use of strategies we and others generated that allow cheap, reliable, and quick assembly of TAL effector constructs with custom repeat arrays (*16*, *19*, *26*, *34*). These strategies are based on "Golden Gate cloning," a method that uses restriction endonucleases that cleave outside their recognition sites to create unique 4-bp overhangs (sticky ends), enabling multiple DNA fragments to be joined in an ordered fashion in a single reaction (*35*, *36*). Vectors for assembling arrays into their native context, in TALENs or as fusions to other protein domains, are available (*16*). Assembly of constructs encoding as many as 31 RVDs can be carried out in about 1 week.

What's on the Horizon for TAL Effectors?

The three-dimensional structure of a TAL effector bound to its DNA target is being pursued by several groups. Such a structure will explain the biophysical nature of RVD-nucleotide preferences and improve our ability to predict mismatch effects on DNA binding. It is also likely to help clarify the relation of array length and composition to affinity and specificity. Not least, it will precisely define the protein/DNA interface so that the architecture of TALENs and other TAL effector fusion proteins can be optimized.

In addition to structural characterization, further development of delivery methods to maximize targeting efficiency in different experimental systems will be important, particularly for DNA editing, which also requires the timed delivery of a repair template. Another issue to be resolved is whether TAL effector binding is affected by epigenetic marks, a question largely unaddressed for zinc finger proteins and homing endonucleases as well. Related to that, whether TAL effector fusions to enzymes such as histone deacetylases or DNA methyltransferases will be effective at modifying epigenetic status will be of interest. Other conceivable uses of TAL effectors, including fusion to cytosine deaminases for site-specific mutagenesis without DNA cleavage, for example, also must be investigated.

Along with the biological questions, legal, sociological, and ethical questions will need to be answered. How should the precise modification of genome sequences or gene expression in living organisms be regulated? Will public perceptions of crop plants and livestock modified through DNA targeting differ from perceptions of conventional genetically modified organisms? What types of modifications under what circumstances and at what levels of risk are acceptable in humans? What about experimental animals or pets? These questions have been relevant since the advent of engineered zinc finger and other DNA targeting proteins, but given the relative ease of customizing TAL effectors and the evidence for their broad targeting range and stringent specificity, it seems they are now more urgent. Though

daunting, these questions speak to the exciting possibilities on the horizon for DNA targeting with TAL effectors.

References and Notes

1. F. D. Urnov, E. J. Rebar, *Biochem. Pharmacol.* **64**, 919 (2002).
2. D. Carroll, *Genetics* **188**, 773 (2011).
3. M. L. Maeder *et al.*, *Mol. Cell* **31**, 294 (2008).
4. J. D. Sander *et al.*, *Nat. Methods* **8**, 67 (2011).
5. H. J. Kim, H. J. Lee, H. Kim, S. W. Cho, J. S. Kim, *Genome Res.* **19**, 1279 (2009).
6. J. Boch *et al.*, *Science* **326**, 1509 (2009); 10.1126/science.1178811.
7. M. J. Moscou, A. J. Bogdanove, *Science* **326**, 1501 (2009); 10.1126/science.1178817.
8. S. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, *Science* **318**, 648 (2007).
9. P. Römer *et al.*, *Science* **318**, 645 (2007).
10. A. J. Bogdanove, S. Schornack, T. Lahaye, *Curr. Opin. Plant Biol.* **13**, 394 (2010).
11. J. Boch, U. Bonas, *Annu. Rev. Phytopathol.* **48**, 419 (2010).
12. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; G, Gly; H, His; I, Ile; K, Lys; N, Asn; and S, Ser.
13. P. Römer, S. Recht, T. Lahaye, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20526 (2009).
14. P. Römer *et al.*, *New Phytol.* **187**, 1048 (2010).
15. G. Antony *et al.*, *Plant Cell* **22**, 3864 (2010).
16. T. Cermak *et al.*, *Nucleic Acids Res.* **39**, e82 (2011).
17. M. Christian *et al.*, *Genetics* **186**, 757 (2010).
18. J. C. Miller *et al.*, *Nat. Biotechnol.* **29**, 143 (2011).
19. F. Zhang *et al.*, *Nat. Biotechnol.* **29**, 149 (2011).
20. M. M. Mahfouz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2623 (2011).
21. R. Morbitzer, J. Elsaesser, J. Hausner, T. Lahaye, *Nucleic Acids Res.* **39**, 5790 (2011).
22. R. Morbitzer, P. Römer, J. Boch, T. Lahaye, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21617 (2010).
23. T. Li *et al.*, *Nucleic Acids Res.* **39**, 359 (2011).
24. B. L. Stoddard, *Structure* **19**, 7 (2011).
25. D. Hockemeyer *et al.*, *Nat. Biotechnol.* **29**, 731 (2011).
26. T. Li *et al.*, *Nucleic Acids Res.* **39**, 6315 (2011).
27. A. J. Wood *et al.*, *Science* **333**, 307 (2011); 10.1126/science.1207773.
28. J. D. Sander *et al.*, *Nat. Biotechnol.* **29**, 697 (2011).
29. P. Huang *et al.*, *Nat. Biotechnol.* **29**, 699 (2011).
30. L. Tesson *et al.*, *Nat. Biotechnol.* **29**, 695 (2011).
31. S. Hacein-Bey-Abina *et al.*, *Science* **302**, 415 (2003).
32. C. Mussolino *et al.*, *Nucleic Acids Res.*; published online 3 August 2011 (10.1093/nar/gkr597).
33. E. M. Händel, S. Alwin, T. Cathomen, *Mol. Ther.* **17**, 104 (2008).
34. R. Geiβler *et al.*, *PLoS ONE* **6**, e19509 (2011).
35. C. Engler, R. Gruetzner, R. Kandzia, S. Marillonnet, *PLoS ONE* **4**, e5553 (2009).
36. C. Engler, R. Kandzia, S. Marillonnet, *PLoS ONE* **3**, e3647 (2008).
37. B. B. Wulff, D. M. Horvath, E. R. Ward, *Curr. Opin. Plant Biol.* **14**, 468 (2011).

10.1126/science.1204094

Science

TAL Effectors: Customizable Proteins for DNA Targeting

Adam J. Bogdanove and Daniel F. Voytas

ARTICLE TOOLS	http://science.sciencemag.org/content/333/6051/1843
RELATED CONTENT	http://science.sciencemag.org/content/sci/333/6051/1797.2.full
REFERENCES	This article cites 36 articles, 12 of which you can access for free http://science.sciencemag.org/content/333/6051/1843#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service