TAL effectors: function, structure, engineering and applications

Amanda Nga-Sze Mak[1], Philip Bradley[2], Adam J Bogdanove[3] and
Barry L Stoddard[1]

TAL effectors are proteins secreted by bacterial pathogens into
plant cells, where they enter the nucleus and activate
expression of individual genes. TAL effectors display a modular
architecture that includes a central DNA-binding region
comprising a tandem array of nearly identical repeats that are
almost all 34 residues long. Residue number 13 in each TAL
repeat (one of two consecutive polymorphic amino acids that
are termed 'repeat variable diresidues', or 'RVDs') specifies the
identity of a single base; collectively the sequential repeats and
their RVDs dictate the recognition of sequential bases along
one of the two DNA strands. The modular architecture of TAL
effectors has facilitated their extremely rapid development and
application as artificial gene targeting reagents, particularly in
the form of site-specific nucleases. Recent crystallographic
and biochemical analyses of TAL effectors have established
the structural basis of their DNA recognition properties and
provide clear directions for future research.

Addresses
[1] Division of Basic Sciences, Fred Hutchinson Cancer Research Center,
1100 Fairview Ave. N. A3-025, Seattle, WA 98109, United States
[2] Division of Public Health Sciences, Fred Hutchinson Cancer Research
Center, 1100 Fairview Ave. N. M1-B514, Seattle, WA 98109, United
States
[3] Department of Plant Pathology and Plant-Microbe Biology, Cornell
University, 334 Plant Science, Ithaca, NY 14853, United States

Corresponding author: Stoddard,
Barry L (bstoddar@fhcrc.org)

Historical background

TAL effectors are trans-kingdom transcription factors
that are secreted by plant pathogenic bacteria in the
genus Xanthomonas [1,2]. Diseases caused by the many
species and pathovars of *Xanthomonas* collectively affect a
wide variety of plants, including several major crop and
ornamental species [3], and their TAL effectors play
critical roles in determining whether the bacterium is
able to infect its host. The first TAL effector identified
was AvrBs3 from *Xanthomonas campestris pv. vesicatoria*
(a pathogen of pepper). AvrBs3 triggers a plant immune
response in strains of pepper that carry the disease resist-
ance gene *Bs3*. First characterized genetically, AvrBs3
activity was shown to correspond to a DNA fragment on a
self-transmissable plasmid that encoded a 125 kilodalton
protein on one strand and an 82 kilodalton protein on the
opposite strand [4]. A comparison of the two 'mirror'
reading frames led to the observation that 'a remarkable
feature of both ORFs is the presence of 17 direct 102 bp
repeats which [within each ORF] share 91% to 100%
homology with each other' [4]. It was subsequently
demonstrated that the open reading frame on the first
strand was the *avrBs3* gene, providing the archetypal
amino acid sequence for this protein class [5].

Subsequent studies revealed members of the *avrBs3*
family in a variety of *Xanthomonas* species, including sev-
eral that like *avrBs3* act as avirulence factors corresponding
specifically to different host resistance genes [6,7], others
that contribute to the pathogen's ability to cause disease in
susceptible plants [8], and some that can play either role
depending on whether the plant carries the corresponding
resistance gene [9,10]. Members of this protein family are
also broadly distributed among diverse isolates of the plant
pathogenic bacterium *Ralstonia solanacearum*, though these
are not yet well characterized [11].

The first clue to the mechanism of TAL effector function
came from the observation that the proteins contain
functional nuclear localization signals (NLS), shown
using a reporter fusion transiently expressed in onion
epidermal cells [12]. Shortly thereafter, localization of
AvrBs3 itself to the nucleus following delivery by the
pathogen during infection was observed and shown to be
required for triggering host immunity [13]. Identification
of a C-terminal acidic activation domain in the members
of the AvrBs3 family, demonstration of the functionality
of this domain in a yeast assay, and discovery of the ability
of the proteins to bind DNA [14–16] provided further
clues regarding the protein domain architecture and
interactions displayed by these genetic factors
(Figure 1) and gave rise to the moniker 'TAL,' which
stands for 'transcription activator-like' [17].

In 2004, a genetic study indicated that mutations in the
gamma subunit of the general transcription factor IIA
(TFIIA) could confer resistance to *Xanthomonad* infec-
tions, thereby suggesting a possible point of interaction
between TAL effectors and the plant host transcriptional
machinery [18]. Subsequent studies demonstrated that

Figure 1
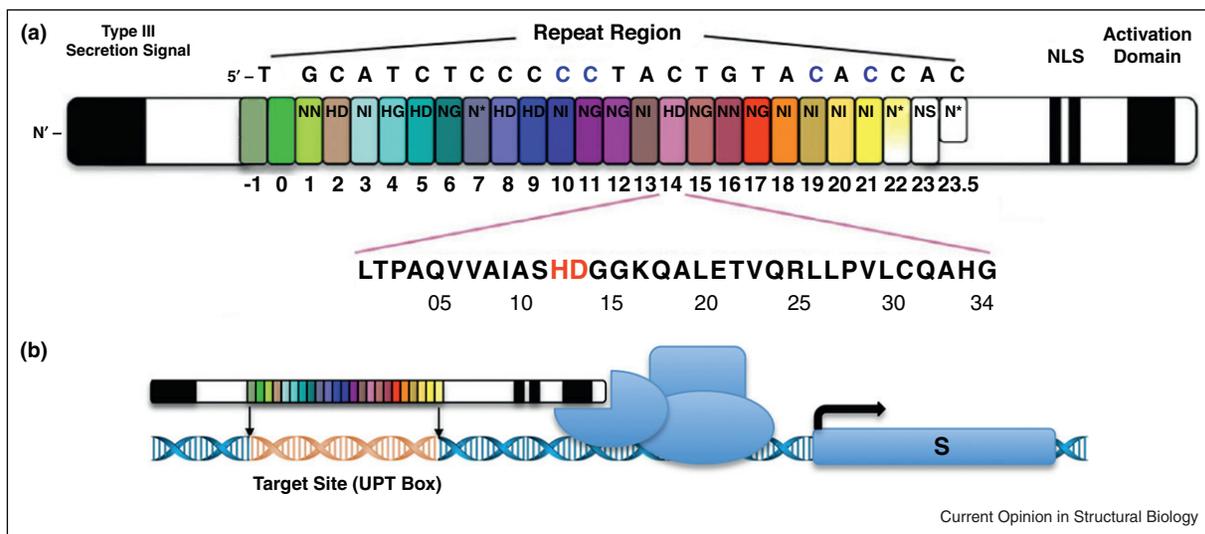
Domain organization and activity of TAL effectors. **(a)** TAL effectors contain N-terminal signals for bacterial type III secretion, variable numbers of tandem repeats that specify the target nucleotide sequence, nuclear localization signals, and a C-terminal region that is required for transcriptional activation. PthXo1 (schematized in this figure) contains 23.5 canonical repeats (color coded to match Figure 2) that contact the DNA target found in the promoter of the rice *Os8N3* gene [20]. Blue bases correspond to positions in the target where the match between protein and DNA differs from the optimal match specified by the recognition code. The sequence of a representative repeat (#14) is shown; RVD residues (HD) that recognize cytosine are red. **(b)** TAL effectors are translocated into the plant nucleus, where they bind to target sites (termed 'UPregulated by TAL' or 'UPT' boxes, or 'EBEs' for 'Effector Binding Elements') that are located in the 5′ promoter regions of genes that are subsequently activated (Termed 'S' for a gene which confers susceptibility to infection as a result of activation, or 'R' for a gene which confers resistance to infection). The C-terminal region of the TAL effector interacts with plant transcriptional machinery as part of the gene activation mechanism. Plants can acquire resistance traits against bacterial infection through at least three separate mechanisms: acquisition of mutations in the EBE that reduce DNA binding affinity, acquisition of mutations in transcription factors that ostensibly inhibit protein–protein association with the TAL effector acidic activation region, or by coupling the sequence of the EBE box to the promoter region of a resistance gene thereby leading to an avirulence phenotype upon infection.

activation of individual plant genes by TAL effectors is linked either to resistance [19] or to susceptibility [20–22] to infection. A pair of reports in 2007 further demonstrated that the avirulence protein AvrBs3 can elicit either a resistance phenotype in pepper plants via direct transcriptional activation of the '*Bs3*' cognate resistance gene [23•] or a susceptibility phenotype in the same species (in the absence of *Bs3*) by activating several genes, including the cell-size regulator *UPA20* [24•]. This analysis resulted in a description of the 'upregulated by AvrBs3' (*UPT*) box: a nucleotide sequence that is conserved among the promoter regions of all the AvrBs3 target genes and is required for activation by AvrBs3. Together, these studies established unequivocally that TAL effectors are trans-kingdom transcription factors.

Recognition code and initial structural analyses

The number of repeats found in TAL effectors varies from five to over thirty, with an average of roughly 17 [1]. Almost all are 34 amino acids in length, and they vary primarily in the identity of the residues at position 12 and 13 in each repeat, a pair of residues that were termed the 'repeat variable diresidue' or 'RVD'. The repeat region always terminates with an apparently truncated repeat,

containing the first 20 residues (including the RVD), which is commonly referred to as a 'half repeat'. Overall, at least two dozen unique RVDs are observed across the known TAL effectors, out of which seven sequences are most common — HD, NG, HG, NN, NS, NI and 'N*'. N* corresponds to a 33 residue repeat with a missing residue within the RVD loop. Two research groups independently demonstrated, in papers published back-to-back in 2009, that the string of RVDs in a TAL effector defines the length and nucleotide sequence of that effector's DNA target, via a one-to-one correspondence of specific RVDs to specific nucleotides [25••,26••]. For example, the presence of an 'HD' RVD within a repeat corresponds to recognition of a cytosine, whereas an 'NG' or 'HG' RVD corresponds to thymine. The modular nature of this recognition mechanism suggested that it could be exploited as a 'code' to predict TAL effector DNA binding sites, and to create gene-targeting proteins using custom arrays of TAL effector repeats.

Recent structural studies of TAL effectors (Figure 2), published side-by-side in early 2012, provided a clear view of the structural basis for the DNA recognition 'code' described above [27••,28••]. The first structure, of an artificially engineered TAL effector termed

Figure 2

(a)

PthXo1
(DNA bound)

(b)

dHax3
(DNA bound
and unbound)

Current Opinion in Structural Biology
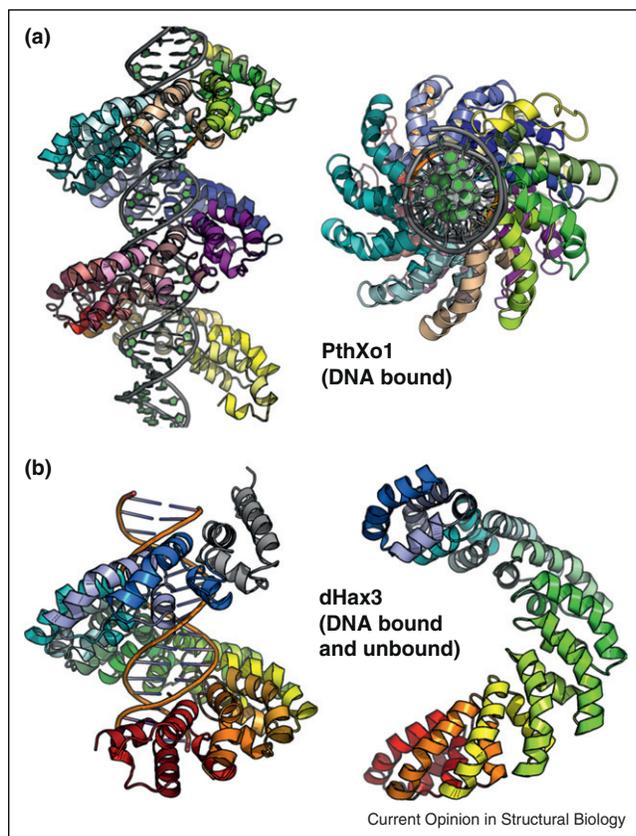
Crystal structures of dHAX3 and pthXo1 TAL effectors. **(a)** The structure of PthXo1 bound to its DNA target site [28••]. The effector contains 22.5 repeat modules each colored separately as shown in Figure 1. The protein–DNA complex is shown from the side of the DNA duplex (left) and looking down the axis of the DNA (right). In the left panel, the N-terminal end of the protein (containing two cryptic repeats that engage the DNA backbone via a series of basic residues, and that also capture the strongly conserved thymine at position 'zero' of the EBE is at the top of the complex (see also Figure 3b)). **(b)** The structure of the artificial dHax3 TAL effector bound to its corresponding DNA target (left) and in the absence of bound DNA (right) [27••]. The protein is displayed with the N-terminal region at the top of both structures. N-terminal cryptic repeat 0, and a truncated portion of repeat −1, is shown as grey ribbon in the left panel. Because a portion of the N-terminal helix of repeat −1 is missing in the construct, the full set of contacts extending to 'thymine zero' in the target site are not entirely formed. Superposition of the DNA bound structures of dHAX3 and PthXo1 yields an overall rmsd for superimposed backbone atoms of approximately 0.9 Å.

'dHAX3', corresponded to a 533 residue construct containing 11 canonical repeats and the half repeat, representing three of the most common RVDs (HD, NG and NS). It was solved in the presence and absence of bound DNA to high resolution (1.85 and 2.4 Å, respectively; PDB entries 3V6T and 3V6P). The second structure, of the naturally occurring TAL effector PthXo1 from *Xanthomonas oryzae* (PDB entry 3UGM), was solved using a high-throughput computational structure prediction and phasing strategy [29•]. Although the PthXo1 structure was

only solved in the presence of bound DNA, and to much lower resolution (d_{min} = 3 Å), it contains over 20 repeats bound to two full turns of DNA and illustrated protein–DNA contacts for six separate types of RVDs (HD, NG, HG, NN, NI, and N*) [28••]. The PthXo1 structure also contains two highly basic 'cryptic' repeats located at the N-terminus that engage the DNA backbone and an essential 5′ thymine residue that immediately precedes the RVD-specified target nucleotides. Because PthXo1 was crystallized in association with its naturally occurring target DNA sequence, the complex contained several examples of RVD-nucleotide mismatches, such as NG versus cytosine, which provided further insight into the structural and biochemical determinants of RVD nucleotide specificity.

The structures both demonstrate that each TAL repeat forms a left-handed, two-helix bundle, in which the two hypervariable residues in each repeat (at positions 12 and 13) are found at the end of the loop that connects the two helices (Figure 3a). The individual repeats carry a relatively neutral overall charge and self-associate to form a right-handed superhelix that wraps around the DNA major groove along the entire length of the DNA target site. The DNA in both structures adopts an unperturbed canonical B-form duplex conformation. The structure of dHAX3 in the absence of DNA indicates that the effector displays a more extended, slightly unwound conformation, although the protein still displays a right-handed superhelical structure with a slightly longer distance separating individual RVDs [27••]. Modeling the conformation of DNA-free dHAX3 around a DNA duplex indicates that a significantly more extended effector conformation might be required for a DNA target search by the unbound protein. That hypothesis agrees with published small angle X-ray scattering (SAXS) data on the full-length PthA TAL effector in the presence and absence of bound DNA, which indicated at least a two-fold reduction in the length of the effector upon target site binding [30].

Recognition mechanism

Both crystal structures also demonstrated that sequence-specific contacts between the effector and the DNA are formed solely by the second residue of each RVD (at position 13 in each repeat) to atoms on the major groove edge of each base on a single contiguous strand of the DNA target. In contrast, the first residue in each RVD (position 12, which is usually occupied by an asparagine or a histidine) serves a largely structural role, forming a hydrogen bond between the side chain and the backbone carbonyl oxygen from position 8 (in the first helix) in each repeat. Those contacts likely help establish a pre-bound conformation of the DNA-contacting RVD loop that ameliorates the substantial entropic cost of binding that would otherwise accompany the ordering of each TAL repeat along the entire length of its target.

Figure 3
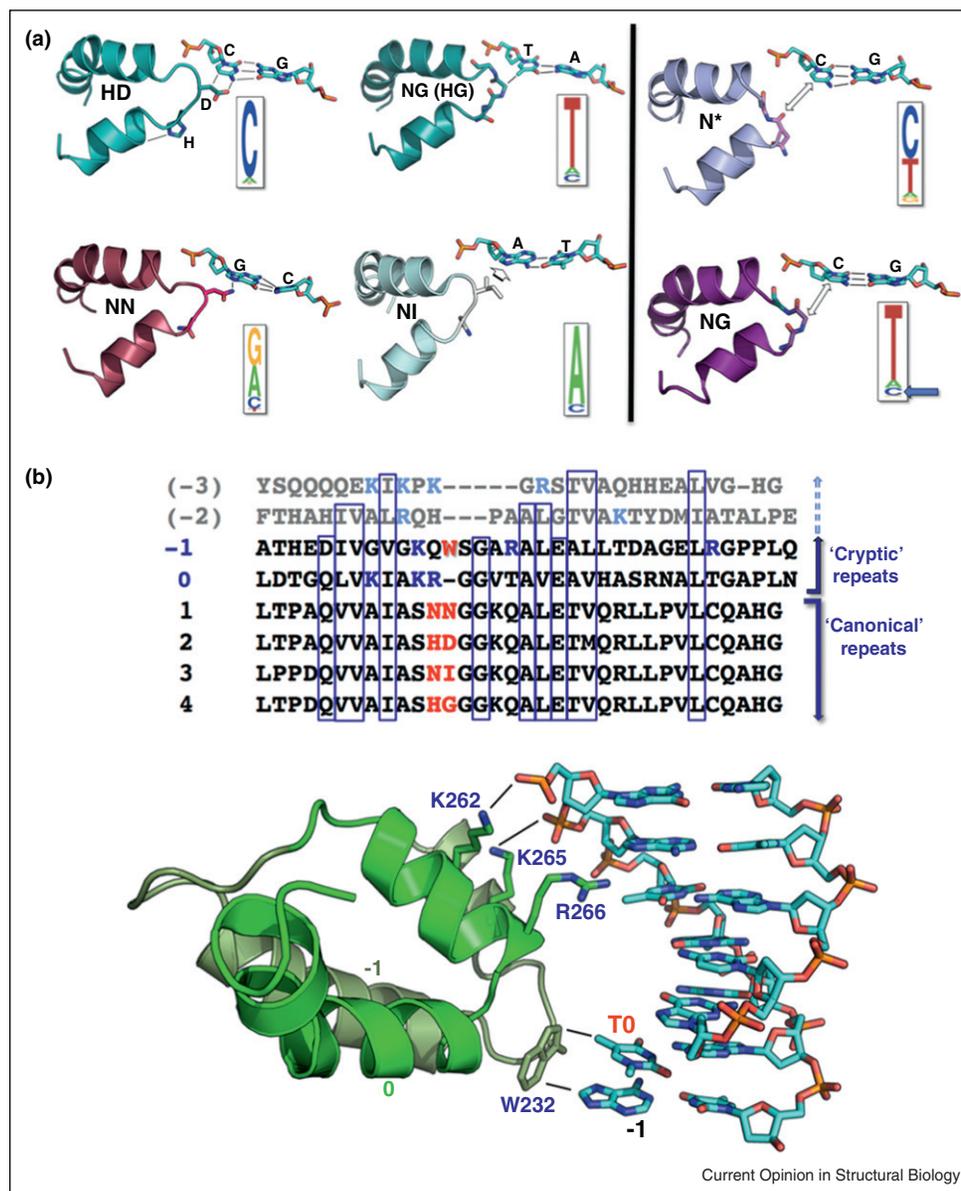
DNA contacts and recognition by TAL effectors. **(a)** Left: interactions formed between four common RVD types (HD, NG, NN and NI) against their cognate nucleotide bases from the PthXo1 target site (cytosine, thymine, purine and adenine, respectively). Whereas the first three RVDs form highly complementary combinations of atomic interactions, NI appears to make desolvating interactions with the neighboring nucleotide and to influence specificity at least in part through steric exclusion, rather than through highly complementary, sequence-specific hydrogen bonds or van der Waals contacts. Recent studies of the overall strength of TAL repeat/RVD interactions to cognate bases indicate that NI is associated with considerably weaker interactions than either HD or NN repeats [33••]. **(b)** Sequence and structural contacts made by residues immediately N-terminal to the central (or 'canonical') TAL repeats in the PthXo1 structure. This region is highly basic (denoted by blue lysine and arginine residues) as compared to the rest of the DNA binding region, and displays limited sequence homology to the C-terminal TAL repeats (indicated by rectangles). The structure of PthXo1 bound to DNA (bottom) demonstrates that residues 220–289 form two cryptic repeats (numbered '0' and '−1') that harbor the same left-handed two-helix bundle as a canonical repeat. These two repeats present a series of basic residues to the DNA backbone (K262, K265 and R266 from the '0' repeat) as well as a single tryptophan residue (W232 from the '−1' repeat) that contacts the 'thymine zero' nucleotide (which is strongly conserved in TAL recognition sites and required for effector activity). The more distal N-terminal region of PthXo1 (and other similar TAL effectors) was not observed in the crystal structure, but appears to contain sequence elements that could form two additional cryptic repeats (numbered '−3' and '−2 and indicated by light grey and light blue fonts and dashed arrow, to distinguish from the region that has been observed in the crystal structure).

The majority of observed contacts between individual RVD residues at position 13 and their corresponding nucleotide bases (Figure 3a) represent interactions that are optimized via either, first, directional hydrogen-bonds for recognition of nucleotide bases (such as HD to cytosine or NN to a purine ring); second, highly complementary packing in the absence of hydrogen bonds (such as between the backbone alpha carbon of a glycine in NG or HG and the extracyclic methyl carbon of a thymine base); or third, interactions that appear to achieve reduced (but not completely negligible) specificity through steric exclusion of alternate bases (in particular, the 'NI' RVD). A fourth type of interaction is represented by the 'N*' RVD, in which truncation of the RVD loop and lack of any side chain at position 13 appears to accommodate any base, presumably with little or no contribution to overall affinity. A recent study has demonstrated that the pattern of contacts and specificity described above can be extended to the recognition of modified bases: the presence of an NG or HG repeat (specific for thymine in an unmodified target) can accommodate similar interactions with a 5-methylcytosine, thus making it possible to identify and/or design potential TAL effectors that can discriminate between target sites that contain methylated CpG sequences and those that are unmodified [31•]. Similarly, an even more recent structural and biochemical study from the same group has demonstrated that TAL effectors can also bind DNA–RNA hybrids, doing so by reading out the DNA strand sequence. Binding of the effector protects such structures from RNase H degradation, and implies that TAL effectors may be used as research tools (or even protein therapeutics) in systems where DNA–RNA hybrids are formed [32•].

A recent recognition study, which examined the function of a variety of artificial TAL effectors that were engineered to contain long strings of single types of repeats and RVDs [33••] indicates that that 'HD' and 'NN' (or 'HN') repeats (which target cytosines and purines, respectively) make the strongest contribution to overall TAL effector function in transcriptional activation assays as compared to the other most common repeat types, whereas effectors containing strings of 'NI' repeats appear to display considerably lower function and reduced specificity. The same study also examined the relative contributions of various other RVD types to TAL effector activity and specificity, and found that an 'NH' RVD (which is found rarely in TAL effectors sequenced to date) demonstrates a strong preference for guanine in the TAL target site. Whether these observations reflect DNA binding properties or other unique requirements for TAL effector activity, and whether they will translate to the function of artificial gene targeting proteins that use the TAL effector repeat scaffold remain to be determined, but such data are clearly of great benefit and have provided important design guidelines for researchers in the field.

An open question regarding TAL effector function is the mechanism by which these highly unusual DNA binding proteins search for and acquire their cognate targets, and the possible role of flanking elements. Possible clues can be found in, first, the relative performance of artificial TAL constructs containing a variety of N-terminal and C-terminal truncations, second, an examination of the sequences immediately N-terminal to the central TAL repeats, and third, the structure of the PthXo1 TAL effector bound to DNA. Various studies of gene targeting proteins constructed using TAL effector scaffolds appear to indicate that the first 120–150 residues are dispensable both for effector and nuclease function, and that further truncations reduce either or both activities [34,35]. The remaining N-terminal region (corresponding roughly to residues 120–254) that immediately precedes the beginning of the canonical TAL repeats contains a highly basic region of the protein, with about 11 conserved arginine and lysine residues that contribute significantly to the overall basic charge of the protein. The structure of the PthXo1 effector bound to its DNA target demonstrated the presence of at least two 'cryptic' repeats (which were termed the 0 and −1 repeats in that model) that form multiple non-specific interactions to the region of the DNA target immediately 5′ of the first sequence specific contact [28] (Figure 3b). This same region of the effector also engages an invariable thymine base (found at position zero of nearly all TAL binding sites) with contacts to the protein backbone and the indole ring of a single tryptophan residue that is found in all *Xanthomonad* TAL effectors. A more recent structural and biophysical analysis of an extended N-terminal region of dHax3 (initially identified via limited proteolytic digests) indicates that as many as four additional cryptic repeats are formed immediately upstream of the central repeat region, and that this region provides the bulk of binding energy required for high affinity target binding and sequence-specific recognition [36•].

Thus, a reasonable model for a TAL effector DNA target search would involve a rapid association and dissociation mechanism that is largely dependent upon the highly basic N-terminal flanking region, followed by a secondary 'annealing' process during each protein–DNA encounter, in which the central TAL repeats (that individually display limited affinity to the negatively charged DNA target) sample the opposing nucleotide base identity sequentially along the target, wrapping around the DNA as long as the cognate sequence is appropriately complementary to each RVD in turn.

Engineering and applications

The biological, bioinformatic and structural studies summarized above have led to an explosion of reports, starting with the initial description of a chimeric TAL effector nuclease in 2010 [37••], that demonstrate the successful creation of a wide variety of gene-targeting reagents using

the TAL effector scaffold, as well as a variety of efficient methods for the rapid creation of such reagents that contain investigator-designed, artificial TAL repeat sequences (recently reviewed in [38–41]). Gene targeting reagents using TAL effector scaffolds have included not only TAL nucleases, but also gene-specific activators and repressors [42–44]. These advances have allowed gene targeting reagents created using TAL effector scaffolds to join and perhaps surpass zinc finger nucleases and homing endonucleases (or 'meganucleases') as commonly employed tools for a variety of genome editing and correction applications. While carefully controlled comparative studies on the performance of each type of reagent on similar targets, and for similar purposes, represent an outstanding area for additional future investigation, a number of recent *in cellulo* and *in vivo* studies suggest that TAL effector-based targeting reagents display robust activity, specificity and low toxicity in a variety of contexts, in addition to their ease of engineering [35,45–52].

Acknowledgements

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- •• of outstanding interest

1. Boch J, Bonas U: **Xanthomonas AvrBs3 family-type III effectors: discovery and function**. *Annu Rev Phytopathol* 2010, **48**:419-436.

2. Bogdanove AJ, Schornack S, Lahaye T: **TAL effectors: finding plant genes for disease and defense**. *Curr Opin Plant Biol* 2010, **13**:394-401.

3. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA *et al.*: **Top 10 plant pathogenic bacteria in molecular plant pathology**. *Mol Plant Pathol* 2012, **13**:614-629.

4. Bonas U, Stall RE, Staskawicz B: **Genetic and structural characterization of the avirulence gene avrbs3 from *Xanthomonas campestris pv. vesicatoria***. *Mol Gen Genet* 1989, **218**:127-136.

5. Knoop V, Staskawicz B, Bonas U: **Expression of the avirulence gene avrBs3 from *Xanthomonas campestris pv. vesicatoria* is not under the control of hrp genes and is independent of plant factors**. *J Bacteriol* 1991, **173**:7142-7150.

6. De Feyter RD, Yang Y, Gabriel DW: **Gene-for-genes interactions between cotton R genes and *Xanthomonas campestris pv. malvacearumavr* genes**. *Mol Plant Microbe Interact* 1993, **6**:225-237.

7. Hopkins CM, White FF, Choi SH, Guo A, Leach JE: **Identification of a family of avirulence genes from *Xanthomonas oryzae pv. oryzae***. *Mol Plant Microbe Interact* 1992, **5**:451-459.

8. Swarup S, De Feyter R, Brlansky RH, Gabriel DN: **A pathogenicity locus from *Xanthomonas citri* enables strains from several pathovars of *Xanthomonas campestris* to elicit canker-like lesions on citrus**. *Phytopathology* 1991, **81**:802-808.

9. Bai J, Choi SH, Ponciano G, Leung H, Leach JE: *Xanthomonasoryzae* pv. *oryzae* avirulence genes contribute differently and specifically to pathogen aggressiveness. *Mol Plant Microbe Interact* 2000, **13**:1322-1329.

10. Swarup S, Yang Y, Kingsley MT, Gabriel DW: **A *Xanthomonas citri* pathogenicity gene *pthA* pleiotropically encodes gratuitous avirulence on nonhosts**. *Mol Plant Microbe Interact* 1992, **5**:204-213.

11. Heuer H, Yin YN, Xue QY, Smalla K, Guo JH: **Repeat domain diversity of avrBs3-like genes in *Ralstonia solanacearum* strains and association with host preferences in the field**. *Appl Environ Microbiol* 2007, **73**:4379-4384.

12. Yang Y, Gabriel DW: ***Xanthomonas* avirulence-pathogenicity gene family encodes functional plant nuclear targeting signals**. *Mol Plant Microbe Interact* 1995, **8**:627-631.

13. Van den Ackerveken G, Marois E, Bonas U: **Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell**. *Cell* 1996, **87**:1307-1316.

14. Yang B, Zhu W, Johnson LB, White FF: **The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein**. *Proc Natl Acad Sci U S A* 2000, **97**:9807-9812.

15. Zhu W, Yang B, Chittoor JM, Johnson LB, White FF: **AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus**. *Mol Plant Microbe Interact* 1998, **11**:824-832.

16. Zhu W, Yang B, Wills N, Johnson LB, White FF: **The C terminus of AvrXa10 can be replaced by the transcriptional activation domain of VP16 from the herpes simplex virus**. *Plant Cell* 1999, **11**:1665-1674.

17. Sugio A, Yang B, Zhu T, White FF: **Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice**. *Proc Natl Acad Sci U S A* 2007, **104**:10720-10725.

18. Iyer AS, McCouch SR: **The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance**. *Mol Plant Microbe Interact* 2004, **17**:1348-1354.

19. Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, Yang F, Chu Z, Wang G-L, White FF *et al.*: ***R* gene expression induced by a type-III effector triggers disease resistance in rice**. *Nature* 2005, **435**:1122-1125.

20. Yang B, Sugio A, White FF: **Os8N3 is a host disease-susceptibility gene for bacterial blight of rice**. *Proc Natl Acad Sci U S A* 2006, **103**:10503-10508.

21. Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, Li X, Fu B, Li Z, Bennetzen JL *et al.*: **Promoter mutations of an essential gene for pollen development result in disease resistance in rice**. *Genes Dev* 2006, **20**:1250-1255.

22. Yang B, Sugio A, White FF: **Avoidance of host recognition by alterations in the repetitive and C-terminal regions of AvrXa7, a type III effector of *Xanthomonas oryzae* pv. *oryzae***. *Mol Plant Microbe Interact* 2005, **18**:142-149.

23. Romer P, Hahn S, Jordan T, Strauss T, Bonas U, Lahaye T: **Plant
• pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene**. *Science* 2007, **318**:645-648.
Together with [24•], this study demonstrates that individual TAL effectors recognize target sites located in the promoter regions of R (resistance) and S (susceptibility) genes in plants in a sequence-specific manner and upregulate the expression of those genes.

24. Kay S, Hahn S, Marois E, Hause G, Bonas U: **A bacterial effector
• acts as a plant transcription factor and induces a cell size regulator**. *Science* 2007, **318**:648-651.
Together with [23•], this paper provides a clear description for the critical activity of a TAL effector as a transcription factor that binds a unique target site (termed a 'UPT box') in a sequence-specific manner; recognition of this sequence within the promoter region of a corresponding plant gene leads to upregulation of that gene that explains the TAL effector-dependent phenotype.

25. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S,
•• Lahaye T, Nickstadt A, Bonas U: **Breaking the code of DNA

binding specificity of TAL-type III effectors. *Science* 2009, **326**:1509-1512.
Together with [26••], this study demonstrates that individual TAL effectors recognize target sites located in the promoter regions of R (resistance) and S (susceptibility) genes in plants in a sequence-specific manner and upregulate the expression of those genes, and describe the correlation between RVD identity in individual TAL repeats and the identity of corresponding nucleotide bases in the DNA target site.

26. Moscou MJ, Bogdanove AJ: **A simple cipher governs DNA**
•• **recognition by TAL effectors**. *Science* 2009, **326**:1501.
Together with [25••], this study described the straightforward correlation between RVD sequences and the DNA target sites sequences that are recognized by TAL effectors.

27. Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N:
•• **Structural basis for sequence-specific recognition of DNA by TAL effectors**. *Science* 2012, **335**:720-723.
This paper describes the high-resolution structures of an artificial, 11.5 repeat-containing TAL effector in the presence and absence of a bound DNA target. The structures demonstrate the structural basis for RVD-dependent sequence recognition by three of the most common RVD types (HD, NG and NS) and illustrates that unbound TAL effectors display a structure that is more elongated than the DNA-bound conformation.

28. Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL:
•• **The crystal structure of TAL effector PthXo1 bound to its DNA target**. *Science* 2012, **335**:716-719.
This paper describes the 3 Å resolution structure of a 22.5-repeat containing, naturally occurring TAL effector (PthXo1) bound to its DNA target from the rice genome. The structure contains six of the most common RVD types (HD, NG, HG, NN, NI and N*) as well as an N-terminal region that contains two 'cryptic' repeats that appear to made additional non-specific contacts to the DNA backbone and a thymine residue at position zero of the target site. Together with [17], this study demonstrates the structural basis of DNA recognition by TAL effectors.

29. Bradley P: **Structural modeling of TAL effector-DNA**
• **interactions**. *Protein Sci* 2012, **21**:471-474.
This methodological paper describes the use of the *Rosetta* protein structure prediction algorithms to systematically model highly constrained symmetric models of possible TAL effector architectures in order to eventually solve the structure of the PthXo1-DNA complex.

30. Murakami MT, Sforca ML, Neves JL, Paiva JH, Domingues MN, Pereira AL, Zeri AC, Benedetti CE: **The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction**. *Proteins* 2010, **78**:3386-3395.

31. Deng D, Yin P, Yan C, Pan X, Gong X, Qi S, Xie T, Mahfouz M,
• Zhu J-K, Yan N *et al.*: **Recognition of methylated DNA by TAL effectors**. *Cell Res* 2012 http://dx.doi.org/10.1038/cr.2012.1127. advance online publication (04-September).
The presence of an NG or HG repeat (specific for thymine in an unmodified target) can accommodate similar interactions with a 5-methylcytosine, thus making it possible to identify and/or design potential TAL effectors that can discriminate between target sites that contain methylated CpG sequences and those that are unmodified.

32. Yin P, Deng D, Yan C, Pan X, Xi JJ, Yan N, Shi Y: **Specific DNA–**
• **RNA hybrid recognition by TAL effectors**. *Cell Rep* 2012 http://dx.doi.org/10.1016/j.celrep.2012.09.001. advance online publication (26-September).
This study demonstrates that TAL effectors can also bind DNA–RNA hybrids, doing so by reading out the DNA strand sequence. Binding of the effector protects such structures from RNase H degradation, and implies that TAL effectors may be used as research tools (or even protein therapeutics) in systems where DNA–RNA hybrids are formed.

33. Streubel J, Blucher C, Landgraf A, Boch J: **TAL effector RVD**
•• **specificities and efficiencies**. *Nat Biotechnol* 2012, **30**:593-595.
This paper provides a series of highly quantitative, systematic comparisons that address the relative affinity and specificity of various RVDs in TAL effectors, and demonstrates that certain RVDs (notably 'NN' and 'HD') appear to engender highly specific and high affinity contacts to their corresponding bases. In addition, the more rarely observed 'NH' RVD appears to confer highly specific recognition of a guanine base within the target site.

34. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ *et al.*: **A TALE nuclease architecture for efficient genome editing**. *Nat Biotechnol* 2011, **29**:143-148.

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

36. Gao H, Wu X, Chai J, Han Z: **Crystal structure of a TALE protein**
• **reveals an extended N-terminal DNA binding region**. *Cell Res* 2012 http://dx.doi.org/10.1038/cr.2012.156. advance online publication (13-November).
This study demonstrates that the N-terminal region in a TAL effector that immediately precedes the central repeats and their RVDs forms an series of extended 'cryptic' repeats that is highly basic in nature and that appears to serve as a 'nucleation region' for DNA binding and subsequent target site recognition.

37. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A,
•• Bogdanove AJ, Voytas DF: **Targeting DNA double-strand breaks with TAL effector nucleases**. *Genetics* 2010, **186**:757-761.
This study provides the first published description of the use of a TAL effector scaffold for targeted gene modification, by coupling its DNA recognition properties to the R.FokI nuclease domain to drive sequence-specific DNA cleavage events that lead to targeted mutagenesis.

38. Bogdanove AJ, Voytas DF: **TAL effectors: customizable proteins for DNA targeting**. *Science* 2011, **333**:1843-1846.

39. Clark KJ, Voytas DF, Ekker SC: **A TALE of two nucleases: gene targeting for the masses?** *Zebrafish* 2011, **8**:147-149.

40. Scholze H, Boch J: **TAL effectors are remote controls for gene activation**. *Curr Opin Microbiol* 2011, **14**:47-53.

41. Perez-Pinera P, Ousterout DG, Gersbach CA: **Advances in targeted genome editing**. *Curr Opin Chem Biol* 2012, **16**:268-277.

42. Geissler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens SE, Boch J: **Transcriptional activators of human genes with programmable DNA-specificity**. *PLoS One* 2011, **6**:e19509.

43. Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P: **Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription**. *Nat Biotechnol* 2011, **29**:149-153.

44. Mahfouz MM, Li L, Piatek M, Fang X, Mansour H, Bangarusamy DK, Zhu JK: **Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein**. *Plant Mol Biol* 2012, **78**:311-321.

45. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC *et al.*: **Genetic engineering of human pluripotent cells using TALE nucleases**. *Nat Biotechnol* 2011, **29**:731-734.

46. Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B: **Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes**. *Nucleic Acids Res* 2011, **39**:6315-6325.

47. Cade L, Reyon D, Hwang WY, Tsai SQ, Patel S, Khayter C, Joung JK, Sander JD, Peterson RT, Yeh JR: **Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs**. *Nucleic Acids Res* 2012, **40**:8001-8010.

48. Garg A, Lohmueller JJ, Silver PA, Armel TZ: **Engineering synthetic TAL effectors with orthogonal target sites**. *Nucleic Acids Res* 2012, **40**:7584-7595.

49. Liu J, Li C, Yu Z, Huang P, Wu H, Wei C, Zhu N, Shen Y, Chen Y, Zhang B *et al.*: **Efficient and specific modifications of the *Drosophila* genome by means of an easy TALEN strategy**. *J Genet Genomics* 2012, **39**:209-215.

50. Moore FE, Reyon D, Sander JD, Martinez SA, Blackburn JS, Khayter C, Ramirez CL, Joung JK, Langenau DM: **Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs)**. *PLoS One* 2012, **7**:e37877.

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

52. Tong C, Huang G, Ashton C, Wu H, Yan H, Ying QL: **Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs**. *J Genet Genomics* 2012, **39**:275-280.