TAL effectors are remote controls for gene activation

Heidi Scholze and Jens Boch

TAL (transcription activator-like) effectors constitute a novel class of DNA-binding proteins with predictable specificity. They are employed by Gram-negative plant-pathogenic bacteria of the genus *Xanthomonas* which translocate a cocktail of different effector proteins via a type III secretion system (T3SS) into plant cells where they serve as virulence determinants. Inside the plant cell, TALs localize to the nucleus, bind to target promoters, and induce expression of plant genes. DNA-binding specificity of TALs is determined by a central domain of tandem repeats. Each repeat confers recognition of one base pair (bp) in the DNA. Rearrangement of repeat modules allows design of proteins with desired DNA-binding specificities. Here, we summarize how TAL specificity is encoded, first structural data and first data on site-specific TAL nucleases.

Introduction

Plant-pathogenic *Xanthomonas* bacteria translocate a cocktail of 30–40 different effector proteins via a type III secretion system (T3SS) into plant cells to manipulate eukaryotic cellular pathways [1]. Members of the TAL (transcription activator-like) effector family include important virulence factors which function as transcription factors for target plant genes [2–6]. This is believed to support bacterial proliferation, but the role of most TALs in virulence is still unknown.

The N-terminal and C-terminal regions are highly conserved in TALs. The N-terminal region contains the secretion and translocation signal for the T3SS, whereas the C-terminal region contains nuclear localization signals and a transcriptional activation domain, both of which are important for protein activity ([2], Figure 1). The most prominent feature of TALs is their central domain that consists of tandem nearly identical repeats. Each repeat is typically 34 amino acids (aa) long, but also repeat types of 30–42 aa can be found ([2], Figure 2). The last repeat is only a half repeat. Repeat-to-repeat variations are limited to a few aa positions including two hypervariable residues at positions 12 and 13 per repeat. Typically, TALs differ in their number of repeats while most contain 15.5–19.5 repeats [2]. The repeat domain determines the specificity of TALs which is mediated by selective DNA binding [7,8]. TAL repeats constitute a novel type of DNA-binding domain [7] distinct from classical zinc finger, helix–turn–helix, and leucine zipper motifs.

The recent understanding of the DNA-recognition specificity of TALs [9,10] has brought TAL studies to a new level, in regard to both functional studies and applied science. In this review we summarize the DNA-recognition code of TALs, novel structural insights, and progress in biotechnological applications. As TAL virulence targets in plants and plant resistance against TALs were recently reviewed [2,3,5,6] these aspects will not be discussed here.

**TALs function as transcription factors**

**TAL frequency**

Proteins with similarity to TALs have only been identified in the plant-pathogenic bacteria *Xanthomonas* spp. and *Ralstonia solanacearum*. The presence, number, and importance of TALs vary in different strains of *Xanthomonas*. Whereas some *Xanthomonas* pathovars carry none or only few TAL genes, others contain up to 28 TAL genes (including pseudogenes) per strain (Table 1). A few TALs have been shown to be essential for virulence, but the contribution of others is minor or not detectable. It is unclear whether TALs with no effect are first, nonfunctional, but possibly serve as recombinatory reservoir, second, only functional in specific host plants, or third, fulfill a task in virulence that has not been detected so far.

Many *R. solanacearum* strains carry homologs of *Xanthomonas* TAL genes (Table 1; [11,12]). However, the N-terminal and C-terminal regions of *R. solanacearum* TALs show low similarity to *Xanthomonas* TALs, and the repeats differ mainly in their C-terminal part (Figure 2). Because of their similar overall repeat-architecture, it is possible that *R. solanacearum* TALs bind to nucleic acids, too.

**DNA-recognition specificity of TALs**

TAL-specific effects are due to specific sequences localized in promoter regions of induced plant genes [7,8,13,14,15]. Characterization of different AvrBs3-induced genes in pepper revealed the presence of a
common promoter element, termed UPA (upregulated by AvrBs3) box [7**,8**,16], which was essential for AvrBs3 recognition and whose function was sensitive to mutations [7**,8**,17,18].

The length of the UPA box roughly matches the number of repeats in AvrBs3, leading to a model to explain AvrBs3 DNA-recognition specificity [9**,10**]. Accordingly, one TAL repeat binds to one DNA base pair (bp), and the specificity of individual repeats is encoded in the hypervariable residues at positions 12 and 13 in each repeat (Figure 2; [9**,10**]; also termed RVD, repeat-variable residue). Some repeat types are specific for a particular DNA bp whereas others recognize more than one bp (9**,10**); Figure 1). Repeat positions other than the hypervariable residues do not exhibit a significant effect on repeat specificity [19]. The target DNA elements of TALs have been termed differently, that is, UPA box [7**], UPT box (upregulated by TAL; [20]) or simply TAL box [9**]. We have suggested to name this element TAL box (e.g. AvrBs3 box, Hax3 box; [9**]), because the DNA sequence controls recognition and binding of the TAL, but not necessarily upregulation of gene expression. TAL boxes all start with a ‘T’ indicating that the N-terminal region preceding the first repeat (termed repeat ‘zero’) contributes to DNA binding (9**,21; Figure 2). Furthermore, it was shown on the basis of artificial TALs containing 1.5–16.5 repeats that a minimum of 6.5 repeats is necessary to induce transcription and 10.5 repeats are sufficient for full induction [9**].

The simple and modular way of how TALs bind to DNA has far-reaching consequences. DNA-binding specificities of TALs can be predicted and repeat modules can be rearranged to generate artificial TALs with novel and predictable DNA-binding specificities [9**,22].

TAL repeat structure
PSIPRED predictions revealed a strong similarity of TAL repeats with tetratrico peptide repeat (TPR) proteins and pentatrico peptide repeat (PPR) proteins [23**,24]. TPRs are 34 aa long, contain two alpha helices, and mediate protein–protein interactions in prokaryotes and eukaryotes (e.g. 79 TPR proteins in Arabidopsis thaliana, 260 TPR proteins in humans) [25]. TPR proteins contain 3–16 degenerated tandem repeats and are involved in diverse cellular processes [25]. In contrast, PPR proteins carry 2–26 either degenerated tandem 35 aa-repeats or tandem alternating 31 aa-repeats, 36 aa-repeats, and 35 aa-repeats [26,27]. PPR proteins mediate RNA binding and are especially abundant in higher plants (e.g. >450 PPR proteins in A. thaliana, six PPR proteins in humans) [26]. As no PPR-3D structures are known TAL repeats were modeled onto TPR protein structures [23**,24]. In such models, the specificity-determining aa residues 12 and 13 of each TAL repeat are located next to another, oriented to the inner face of the superhelix [23**,24] which fits well to a possible TAL–DNA interaction (Figure 2). Typically, protein repeat units (e.g. TPR and PPR) are highly degenerate and the consensus structure only forms a scaffold to present functional residues for interactions [25,28]. In contrast, TAL repeats show an extraordinarily high degree of repeat-to-repeat identity. This probably generates a highly symmetric framework to position the two specificity-determining hypervariable residues to its highly symmetrical binding partner, the DNA double helix.

Nuclear magnetic resonance (NMR) structural data of a 1.5 TAL repeat polypeptide showed that the TAL repeat
folds into a helix–turn–helix–turn structure reminiscent of TPRs (Figure 2). The hypervariable residues were located at the end of the first helix. The two alpha helices of one repeat are arranged in antiparallel orientation and likely interact with each other via hydrophobic residues [23]. Possibly, the structure of the whole TAL repeat domain differs, because of repeat-to-repeat interactions which could not be resolved in the short 1.5 repeat fragment.

**TAL–DNA interaction**

TAL repeats specifically bind to DNA *in vitro* in a sequence-dependent manner suggesting that no eukaryotic host factor is needed for DNA binding and specificity [7,8,9,17,18,21]. It is suggestive that the hypervariable residues, which determine the specificity of each repeat directly interact with the DNA bases.

Generally, repeat-containing proteins carry tandem arrays of a short structural motif and the repetitive structure typically forms an extended interaction interface to bind diverse ligands [28]. The repetitive units can be arranged similarly to a coil spring in a so-called solenoid structure [29]. The vast majority of solenoids in nature are right-handed (i.e. a right-handed coil). In contrast, the Zurdo ('left-handed' in Spanish) domain has recently been described as a left-handed alpha-helical solenoid that binds to DNA [30]. Often, solenoid units are stacked slightly shifted to each other resulting in a twisted right-handed or left-handed continuous superhelix [29]. Based on predictions [23,24] and to accommodate the DNA, TAL repeats likely twist in a right-handed superhelix around the right-handed DNA helix. A manual fit of possible TAL–DNA complexes shows that right-handed or left-handed TAL repeat solenoids might wrap around the DNA in a sunflower-like shape (Figure 2).

TAL effectors with 17.5 repeats, for example, AvrBs3, interact with a target DNA sequence (one repeat per bp plus flanking sequences [9]) that corresponds to nearly two helical turns. The predicted right-handed superhelical TAL repeat structure can be expected to be flexible to wrap around the DNA double helix (Figure 2). The modeled structure of the TAL repeat region [23] is too large (approximately twice the size) to accomplish interaction of one repeat to one bp as required from its DNA specificity [9,10]. Interestingly, the TAL repeats undergo conformational changes upon DNA interaction to a more compact structure and the helical content of the repeats is decreased upon DNA binding [23]. This suggests that TAL–DNA complex formation is a dynamic process that involves repeat and/or DNA conformational changes.

**Biotechnology aspects**

**TALs as transcriptional remote controls**

TALs can be tailored to target specific DNA sequences and, accordingly, specifically induce target reporter genes as well as endogenes in plants [9,19,22]. The position of a TAL box defines the TAL-dependent transcriptional start site at approximately 40–60 bp downstream of the box [7,17,18,20]. Obviously, TAL effectors control the transcriptional start site analogous to the TATA box binding protein which may indicate similar functions. In addition, different TAL boxes can be combined in a *cis*-regulatory element in one promoter to render the downstream gene inducible by more than one TAL [20]. This has consequences for biotechnology, because a gene of interest can be induced using different TALs and these TALs can target DNA boxes at different places of the promoter region.

**TAL technology versus zinc finger technology**

The programmable DNA-binding specificity of TALs enables biotechnological applications comparatively to the zinc finger (ZF) technology [31]. One ZF domain binds to 3–4 DNA bp, and the specificity of ZF domains can be modified. A tandem array of 3–6 ZFs corresponds to target sites of 9–18 bp which is sufficient to target unique sites in complex genomes. The ZF technology has been used to generate artificial transcription factors, repressors, methylases, recombinases, and ZF nucleases (ZFNs) [31]. The binding specificity of a ZF array is not completely predictable, because specificities of neighboring ZFs depend which results in highly laborious screening of libraries to identify suitable candidates [32,33]. In contrast, TALs have an obvious advantage, because the TAL–DNA-binding specificity is unambiguously predictable and TAL repeat specificity is obviously neighbor-independent [9,10].

**TAL nucleases promote genome editing**

Zinc finger nucleases (ZFNs) are fusion proteins of tandem ZF domains and a nonspecific endonuclease, typically the C-terminal nuclease domain of the type II restriction enzyme FokI [34–37]. FokI enzymatic activity depends on dimerization after DNA binding and, accordingly, two ZFNs are designed to bind to adjacent DNA sequences to allow FokI dimerization and DNA cleavage within the spacer region between both binding sites [36; Figure 3]. DNA double-strand breaks are generally repaired by one of two alternative mechanisms, nonhomologous end-joining (NHEJ) or homologous DNA recombination (HDR) [36]. The first mechanism is typically accompanied by mutagenic deletions or insertions whereas the latter can also mediate recombination between endogenous sequences and exogenously supplied DNA fragments. Targeted cuts enable targeted DNA integration, gene replacement or gene knockout in complex genomes [34–37]. Accordingly, the ZFN technology has been used for genome modification of yeast, plants, mammals, and humans [36].

In addition to ZFNs, chimera between TALs and the FokI nuclease have recently been constructed and
Possible TAL effector repeat arrangements. (a) NMR structure of 1.5 repeats of the TAL effector PthA2 (2KQ5) forming three α-helices [23*]. (b) Cartoon of three consecutive repeats forming a predicted right-handed superhelix of repeats twisted 36° to fit the right-handed DNA helix. Manual fit of 10 individual repeats onto DNA (black). To connect individual helices, the long alpha helices would have to be kinked. Top and bottom models are based on left-handed and right-handed solenoids (repeat twist), respectively. Repeats of the right-handed solenoids are shown tilted. α-Helical
TAL nucleases (TALNs) promote genome editing. (a) TALNs are fusions between TAL effectors and the FokI endonuclease domain. A tailored TAL repeat domain controls DNA-binding specificity. (b) Two TALNs bind neighboring DNA boxes and FokI dimerization induces DNA cleavage in the spacer region between the boxes. DNA double-strand breaks can promote nonhomologous end-joining (NHEJ) or homologous DNA recombination (HDR) enabling targeted genome modifications like deletions or insertions.

Table 1

<table>
<thead>
<tr>
<th>Host</th>
<th>TALs a</th>
<th>Reference b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xanthomonas spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X. campestris pv. musacearum</td>
<td>Banana</td>
<td>0 ACHT01000000</td>
</tr>
<tr>
<td>X. campestris pv. vasculorum</td>
<td>Sugarcane</td>
<td>0 ACHS01000000</td>
</tr>
<tr>
<td>X. campestris pv. vesicatoria</td>
<td>Pepper and tomato</td>
<td>0-1 AM039952</td>
</tr>
<tr>
<td>X. gardneri</td>
<td>Pepper and tomato</td>
<td>0-1 AM920689; CP000050; AE008922</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>Brassicaceae</td>
<td>0-1 AY780631</td>
</tr>
<tr>
<td>X. axonopodis pv. glycines</td>
<td>Soybean</td>
<td>1 [44]; [46]; <a href="http://www.cmr.jcvi.org">www.cmr.jcvi.org</a></td>
</tr>
<tr>
<td>X. axonopodis pv. manihotis</td>
<td>Cassava</td>
<td>≥1 [45]; [46]</td>
</tr>
<tr>
<td>X. campestris pv.armoraciae</td>
<td>Brassicaceae</td>
<td>0-3 [46]; [46]; <a href="http://www.cmr.jcvi.org">www.cmr.jcvi.org</a></td>
</tr>
<tr>
<td>X. axonopodis pv. citri</td>
<td>Citrus</td>
<td>2-4 [47-49]; AE008923</td>
</tr>
<tr>
<td>X. fuscans subsp. aurantifolii</td>
<td>Citrus</td>
<td>1-4 [47-49]; AE008923</td>
</tr>
<tr>
<td>X. translucens pvs.</td>
<td>Cereals and grasses</td>
<td>0-6 [50]</td>
</tr>
<tr>
<td>X. campestris pv. malvacearum</td>
<td>Cotton</td>
<td>6-10 [51,52]</td>
</tr>
<tr>
<td>X. oryzae pv. oryzae</td>
<td>Rice</td>
<td>7-16 [50]; CP000967; AP008229; AE013598</td>
</tr>
<tr>
<td>X. oryzae pv. oryzicola</td>
<td>Rice</td>
<td>12-28 [50]; AAQN01000001</td>
</tr>
<tr>
<td>Ralstonia spp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. solanacearum</td>
<td>Broad host range</td>
<td>0-1 [11]; AL646052</td>
</tr>
</tbody>
</table>

a The number of TAL effectors per strain was determined by Southern blot or genomic sequence. Southern blots did not reveal pseudogenes and might not have identified all alleles. Further information on Xanthomonas genomes and TALs can be found at www.xanthomonas.org and www.cmr.jcvi.org.

b GenBank accession numbers are given as reference for sequenced genomes and plasmids.

c The probe for Southern analysis yielded several bands, but contained TAL and TAL-flanking sequences which precludes allele number determination.

TAL nucleases (TALNs) are fusions between TAL effectors and the FokI endonuclease domain. A tailored TAL repeat domain controls DNA-binding specificity. Two TALNs bind neighboring DNA boxes and FokI dimerization induces DNA cleavage in the spacer region between the boxes. DNA double-strand breaks can promote nonhomologous end-joining (NHEJ) or homologous DNA recombination (HDR) enabling targeted genome modifications like deletions or insertions.

TAL nucleases (TALNs or TALENs; Figure 3: [38*,39*]). Different TALs were combined with a FokI nuclease domain at the N-terminus or C-terminus. Expression of TALN pairs that bind to adjacent DNA boxes promoted DNA cleavage, followed by targeted HDR in yeast with comparable efficiency as ZFNs [38*,39*]. Similarly, TALNs with artificial arrangements of repeats were designed to target a given DNA sequence...
Surprisingly, some artificial TALNs showed no activity, indicating that some unknown parameters still govern efficiency of TALNs [38*]. The ZFN technology has been hampered by cytotoxicity possibly due to off-target cleavage at multiple genomic sites [40,41]. In contrast, TALs are highly sensitive to changes in their target DNA boxes and three to four mutations typically block recognition [9**,19] which indicates that off-targets might be limited.

Conclusions

Rarely has a protein in nature been identified whose specificity can be readily predicted from its amino acid sequence and furthermore, easily modified. The simple structure-function encryption of the TAL–DNA-binding specificity is both elegant and intriguing. It immediately suggests use of TAL effectors as programmable DNA-targeting devices. First results showed that TAL nucleases have specific activity and certainly other applications will be explored, soon. TAL proteins have the potential to rival the existing ZF technology and establish a game-changing variety of novel DNA sequence-specific applications (specific gene activators, repressors, DNA-probes, gene replacement, and mutations) and hopefully therapeutics.

Note added in proof

Recently, TAL nucleases were successfully employed to edit human chromosomal targets and designed TALs induced expression of endogenous human genes [53].

Acknowledgements

We thank D. Büttner, R. Koebnik, and U. Bonas for discussions and helpful suggestions on the manuscript. Research in our laboratory is supported by a grant from the Deutsche Forschungsgemeinschaft (SPP 1212) to J.B.

References and recommended reading

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

● of special interest

●● of outstanding interest


First report (with Ref. [8**]) that TAL effectors bind to a specific DNA sequence in promoters of target plant genes. AvrBs3 induces hypotrophy via expression of UPA2O, a cell-size regulator from pepper.


First report (with Ref. [7**]) that TAL effectors bind to a specific DNA sequence in promoters of target induced plant genes. AvrBs3 induces expression of the pepper resistance gene Bs3 leading to cell death. Bs3 encodes a putative flavin-monoxygenase.


Experimental evidence that solved the code of DNA recognition specificity of TAL effectors. First artificial TALs were generated and shown to exhibit the predicted specificity. See also Ref. [10**].


11. Heuer H, Yin Y-N, Xue Q-Y, Smalla K, Guo J-H: Bioinformatic analysis that supported a model for DNA recognition specificity of TAL effectors. See also Ref. [9**].


First report that plant resistance genes against TAL effectors can be transcriptionally induced by TAL effectors. The rice resistance gene Xa27 and the cognate TAL AvrXa27 were cloned.


A combination of different TAL boxes upstream of a gene renders this gene inducible by several TALs. Identification of the AvrXa27 target box.

genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, Xanthomonas oryzae pv. oryzae. New Phytol 2010, 187:1048-1057.


34. Davis D, Stokoe D: Zinc finger nucleases as tools to understand and treat human diseases. BMC Med 2010, 8:42.


