TAL effectors are important virulence factors of bacterial plant pathogenic Xanthomonas, which infect a wide variety of plants including valuable crops like pepper, rice and citrus. TAL proteins are translocated via the bacterial type III secretion system into host cells and induce transcription of plant genes by binding to target gene promoters. Members of the TAL effector family differ mainly in their central domain of tandemly arranged repeats of typically 34 amino acids each with hypervariable di-amino acids at positions 12 and 13. We recently showed that target DNA-recognition specificity of TAL effectors is encoded in a modular and clearly predictable mode. The repeats of TAL effectors feature a surprising one-repeat-to-one-bp correlation with different repeat types exhibiting a different DNA base pair specificity. Accordingly, we predicted DNA specificities of TAL effectors and generated artificial TAL proteins with novel DNA recognition specificities. We describe here novel artificial TALs and discuss implications for the DNA recognition specificity. The unique TAL-DNA binding domain allows design of proteins with potentially any given DNA recognition specificity enabling many uses for biotechnology.

The genus Xanthomonas consists of Gram-negative phytopathogenic bacteria that occur worldwide and infect a large spectrum of host plants. Virulence of nearly all Xanthomonas strains relies on a type III secretion system that translocates bacterial virulence proteins (termed effectors) directly into plant cells.1 Xanthomonas TAL (transcription activator-like) effectors function as transcriptional activators of plant genes in the plant nucleus and include major virulence factors.2,3 The specific activity of TAL effectors is based on a central domain of tandemly arranged nearly identical repeats of typically 34 amino acids. Similarity of the last repeat extends only over the first 20 amino acids and is referred to as a half repeat. Repeat to repeat variations occur predominantly at amino acid positions 12 and 13 which have therefore been termed hypervariable.2 Members of the TAL effector family differ mainly in the number and order of their repeats. The archetype TAL effector, AvrBs3, induces transcription of more than 20 pepper genes (termed UPA for upregulated by AvrBs3) by binding to a conserved element (UPA box) in the promoter of target genes.4,7 DNA binding is mediated by the repeat region.4,5,7-9 Host factors that influence binding of TAL effectors to DNA have not been identified, so far. TAL repeats resemble no classic DNA binding domain. How TAL repeats interact with DNA and how TAL effector specificity is encoded has remained a mystery until recently.

We developed a model to account for the DNA-binding specificity of AvrBs3 and other TAL effectors (Fig. 1).10 Our hypothesis was based on the primary observation that AvrBs3 binds to the UPA box.4,7 According to our model, one TAL repeat corresponds to one base pair in the target DNA. The two hypervariable amino acids per repeat provide recognition specificity and allow classification of repeats into different repeat types (e.g., HD-, NG-, NI-, NS-repeats in AvrBs3, Fig. 1A). In total, the sequence of repeats thus binds to a consecutive DNA sequence. We additionally observed that the recognition specificity of AvrBs3 and other

Key words: Xanthomonas, type III secretion, transcriptional activator, AvrBs3, plant pathogen, repeats, RVD, biotechnology

Submitted: 05/12/10
Revised: 06/28/10
Accepted: 06/30/10

Previously published online: www.landesbioscience.com/journals/virulence/article/12863

*Correspondence to: Jens Boch; Email: jens.boch@genetik.uni-halle.de
TAL effectors are extended by one base pair at the 5’ end. Presumably, a region (termed repeat 0) N-terminally adjacent to the repeat region determines a specificity for thymine (sense strand). Therefore, the target box of AvrBs3 (17.5 repeats) is 19 bp long (Fig. 1A). According to our model, we predicted and successfully verified hitherto unknown target specificities of four additional TAL effectors, Hax2, Hax3, Hax4 and AvrXa10. All four TAL proteins specifically recognized the corresponding predicted target DNA sequences and efficiently induced transcription of the uidA (GUS) reporter gene in planta. We further demonstrated that certain repeat types exhibit preference for one specific base pair (NI-, HD- and NG-repeats) or two alternative base pairs (NN-repeats) or are apparently unselective (NS-repeats) (Fig. 1B). In a parallel study, a similar model for TAL recognition specificity was published based on bioinformatic analyses of target induced plant genes. Using bioinformatics, no influence of neighboring repeats on the recognition specificity of a given repeat were found, suggesting that each repeat acts as an independent DNA-binding module. Together, both studies demonstrated that the DNA-recognition specificity of TAL effectors is encoded in a surprisingly simple and modular way, unparalleled to other known DNA-binding domains.

If TAL repeats function independently, it should be possible to rearrange the repeats in the repeat region to generate effectors with novel and predictable DNA-recognition specificities. We generated a collection of artificial TAL effectors (ArtX) with a near-random (first repeat is always NI and last repeat always NG) mixture of four key repeat types (NI, HD, NN, NG) that correspond to A, C, G and T, respectively, in the DNA. Five effectors with 12.5 (ArtX1-5) and five effectors with 10.5 repeats (ArtX6-10) were chosen (Fig. 2A) to deduce target DNA boxes according to the DNA-specificity code (Fig. 1B). ArtX1-ArtX3 have been described before. The target DNA boxes of ArtX1-10 were inserted into the minimal Bs3 promoter which has low basal activity. Gene expression was analyzed in planta following Agrobacterium-mediated cotransformation using GUS as reporter (Fig. 2B). All ten artificial TAL effectors led to a strong induction of promoters containing the corresponding target DNA box (Fig. 2A). The strength of gene induction is comparable in all cases, indicating that the order of repeats and the repeat composition does not control the efficiency of ArtX1-10. The ArtX1-10 activity is similar to the activity of natural TAL effectors (data not shown) although this has not been compared in detail. Earlier, we showed that ArtX1, ArtX2 and ArtX3 specifically recognize only the corresponding DNA sequences, although the ArtX2 and ArtX3 target DNA boxes differ in only four positions emphasizing that DNA recognition by the repeat domain is very specific. All artificial effector constructs were generated with an N-terminal GFP-tag which does not interfere with protein function. GFP fluorescence occurs exclusively in the plant nucleus indicating that full-length fusion proteins are expressed and localize correctly (Fig. 2C).

In 113 published TAL sequences the size of the repeat region differs from 1.5 to 35.5 repeats, but it was unclear how

Figure 1. Model of the TAL effector-DNA binding code. (A) TAL effectors contain a central domain of typically 34-amino acid tandem repeats. The last repeat is shorter (half-repeat). Repeat types are specified by amino acids 12 and 13 (shaded in gray in the sample repeat above and indicated in the 17.5 repeats of AvrBs3). One repeat recognizes one base pair on the DNA. Target DNA boxes (upper, sense strand) of AvrBs3 in the promoters of pepper UPA20 and Bs3 and the optimal DNA box (N = A, C, G, T) according to the repeat specificity are aligned to the AvrBs3 repeats. Similar colors indicate matching repeat-DNA combinations. Black color and grey shading shows mismatches. (B) Repeat types and their DNA base-recognition specificity (upper, sense strand).

Figure 2 (See opposite page). Artificial TAL effectors recognize predicted target boxes. (A) Repeat composition of artificial effectors (ArtX) with amino acids 12 and 13 per repeat. Corresponding target DNA boxes were deduced according to the repeat specificity. The ArtX1-box is shown as an example below ArtX1. GUS reporter constructs carrying ArtX-boxes were coderviled via Agrobacterium into Nicotiana benthamiana with matching 35S-driven artX1-10 and empty T-DNA (ev), respectively (error bars show SD; n = 3 samples). (A) Strawberry, 4-methylumbelliferone. 35S::uidA served as control. Leaf disks were stained with X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide). (B) ArtX-boxes were inserted in front of the minimal tomato BrA promoter into a GUS (uidA) reporter vector. (C) Artificial TAL effectors localize to the plant cell nucleus. All TAL effectors were cloned as N-terminal GFP fusions. Confocal laser scanning microscopy of lower epidermal N. benthamiana cells transformed with 35S::artX6 or 35S::gfp. GFP fluorescence, DAPI (4’,6-diamidino-2-phenylindole) fluorescence and an overlay of both channels is shown.
the repeat number influences DNA binding. TAL effectors with a low number of repeats have less protein-DNA interactions than TAL effectors with many repeats and should thus have lower DNA affinity. Indeed, we could show that a minimum of 6.5 repeats is required to activate a reporter while TAL effectors with 10.5 and more repeats function equally well. Published TAL amino acid sequences reveal 23 different repeat types based on amino acids 12 and 13. Seven repeat types (NI, NG, NN, NS, N*, HD, HG) are by far the most common ones and resemble the dominant repeats. The high number of repeat types might reflect different specificities or DNA interaction capacities (strong/weak binding or strong/weak mismatch penalty). Possibly some repeats may not contribute to binding at all.

The position or nature of a repeat in the repeat region seems to have an influence on DNA binding, because some non-matching repeat-DNA combinations strongly impair DNA recognition, whereas others don’t have an obvious effect. Using the repeat-specificity code, optimal target DNA boxes of known TAL effectors were predicted and potential TAL-target DNA sequences in promoters of plant genes were identified. Three TAL-target DNA boxes in rice gene (Xa13, OiTFX1, Os11N3) promoters were recognized by corresponding TAL effectors (PthXo1, PthXo6, AvrXa7) although they contain up to seven mismatches (AvrXa7/Os11N3) in comparison to the predicted optimal DNA boxes.

It has been speculated that TAL effectors with high numbers of repeats (AvrXa7 has 25.5 repeats) tolerate more mismatches and that binding affinity depends on the sum of matching repeat-DNA combinations. Possibly, though, the interaction dynamics are more complex. AvrBs3 (17.5 repeats) does not efficiently induce gene expression if the natural target box in the pepper Bt3 promoter (contains three repeat-DNA mismatches, Fig. 1A) is modified to contain certain combinations of four mismatches (=14 matching repeats). In addition, AvrBs3 recognizes well the UP4 box in the UP420 promoter (contains two repeat-DNA mismatches, Fig. 1A), but does not efficiently recognize DNA sequences with two additional

mismatches (=14 matching repeats). Similarly, PthXo1 (23.5 repeats) recognizes a target box with four mismatches (repeat 10 NI-C, repeat 11 NG-C, repeat 19 NI-C, repeat 21 NI-C) but not one that contains one additional mismatch (repeat 1 NN-T; =18.5 matching repeats). In contrast, Hax3 (11.5 repeats) and several artificial TAL effectors with 10.5 repeats (Fig. 2A) strongly induce promoter-reporter fusions containing corresponding boxes even though the overall number of matching repeat-DNA combinations is lower.

Therefore, we now postulate that interaction between the TAL repeat region and DNA is influenced by at least four parameters: (1) a minimal number of repeats is needed for efficient TAL-DNA interaction; (2) non-matching repeat-DNA combinations (negative effect) contribute more strongly to overall interaction than matching repeat-DNA combinations (positive effect); (3) different repeat types exhibit different DNA-interaction capabilities (binding strength or mismatch penalty); (4) the effect of individual non-matching repeat-DNA combinations are position- and context-dependent. To illustrate this, the binding dynamics of the TAL repeat region might resemble the interaction of oligonucleotides with single stranded DNA (e.g., primer in a PCR reaction). Here, the interaction is also dependent on total length, number of mismatches, position of the mismatch and neighboring matches/mismatches.

TAL effectors initiate a transcriptional start site about 40–60 bp downstream of the TAL-box, indicating that they assemble a transcriptional initiation complex at a position different from the endogenous one (e.g., the TATA-box). Which host components are part of this complex and which initiation steps are possibly directly performed by the TAL effectors is unknown so far. TAL effectors contain a C-terminal acidic activation domain that is required for full function in planta. The molecular role of this activation domain in TAL-mediated gene induction is still unknown. Recently, it was shown that different TAL-boxes can be combined into one promoter to render this promoter responsive to several TAL effectors. This demonstrated that TAL-boxes can be inserted at different positions into a promoter and promoters can be generated that are induced by several TAL effectors. This might be useful to generate plants carrying resistance genes under control of such promoters. These plants will be resistant to infection of bacteria carrying either one of the corresponding TAL effectors.

Breaking the code of DNA recognition specificity of TAL effectors was a milestone in understanding function of this important effector family. It opened the door to use TAL effectors for biotechnology as both molecular devices to control gene expression and universal DNA binding domains with predictable specificity.

Several points need to be addressed in the future: (1) Which plant genes are activated by TAL effectors and how do these genes affect the outcome of the plant-bacterial interaction? (2) How is transcription initiated? (3) Can TAL effectors be modified to function in humans and other kingdoms of life? (4) Can the repeat region be used in biotechnology for specific DNA-targeting in fusions with functional domains that cut, modify or manipulate DNA? Microbes developed TAL effectors as potent nano tools to manipulate expression in plant cells and we might now use this unique invention for biotechnology.

Acknowledgements
We thank A. Richter for constructs, A. Landgraf for excellent technical assistance and U. Bonas for continuous support. This work was supported by the Deutsche Forschungsgemeinschaft (SPP 1212).

References
1. Büttner D, He SY. Type III protein secretion in plant pathogenic bacteria. Plant Physiol 2009; 150:1656-64.
2. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol 2010; 48: DOI: 10.1146/annurev-phyto-080508-081936; In press.
3. Schornack S, Meyer A, Romer P, Jordan T, Lahaye T. Gene-for-gene-mediated recognition of nucleotide-targeted AvrBs3-like bacterial effector proteins. J Plant Physiol 2006; 163:256-72.
4. Kay S, Hahn S, Marotz E, Haufe G, Bonas U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 2007; 316:648-51.
5. Kay S, Hahn S, Marotz E, Wieduwilt R, Bonas U. Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Mep16. Plant J 2009; 59:859-71.
6. Marois E, Van den Ackerveken G, Bonas U. The Xanthomonas type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. Mol Plant-Microbe Interact 2002; 15:637-46.

7. Römer P, Hahn S, Jordan T, Strauß T, Bonas U, Lahaye T. Plant-pathogen recognition mediated by promoter activation of the pepper Bt3 resistance gene. Science 2007; 318:645-8.

8. Römer P, Recht S, Strauß T, Elsaesser J, Schornack S, Boch J, et al. 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 2010; DOI:10.1111/j.1469-8137.2010.03217.x; In press.

9. Römer P, Strauß T, Hahn S, Scholze H, Morbitzer R, Grau J, et al. Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bt3 alleles. Plant Physiol 2009; 150:1697-712.

10. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 2009; 326:1509-12.

11. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science 2009; 326:1501.

12. Römer P, Recht S, Lahaye T. A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proc Natl Acad Sci USA 2009; 106:20526-31.

13. Szurek B, Marois E, Bonas U, Van den Ackerveken G. Eukaryotic features of the Xanthomonas type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. Plant J 2001; 26:523-34.

14. Zhu W, Yang B, Chittor JM, Johnson LB, White FF. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. Mol Plant-Microbe Interact 1998; 11:824-32.