*This copy is for your personal, non-commercial use only.*

# Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors

Jens Boch,* Heidi Scholze, Sebastian Schornack,† Angelika Landgraf, Simone Hahn, Sabine Kay, Thomas Lahaye, Anja Nickstadt,‡ Ulla Bonas

The pathogenicity of many bacteria depends on the injection of effector proteins via type III secretion into eukaryotic cells in order to manipulate cellular processes. TAL (transcription activator–like) effectors from plant pathogenic *Xanthomonas* are important virulence factors that act as transcriptional activators in the plant cell nucleus, where they directly bind to DNA via a central domain of tandem repeats. Here, we show how target DNA specificity of TAL effectors is encoded. Two hypervariable amino acid residues in each repeat recognize one base pair in the target DNA. Recognition sequences of TAL effectors were predicted and experimentally confirmed. The modular protein architecture enabled the construction of artificial effectors with new specificities. Our study describes the functionality of a distinct type of DNA binding domain and allows the design of DNA binding domains for biotechnology.

P hytopathogenic bacteria of the genus *Xanthomonas* cause severe diseases on many crop plants. Pathogenicity relies on the translocation of effector proteins into the plant cell cytoplasm via the type III secretion system (*1–5*). Members of the large transcription activator-like (TAL) effector family are key virulence factors of *Xanthomonas* (*4–7*) and reprogram host cells by mimicking eukaryotic transcription factors (*8–13*). TAL effector–mediated gene induction leads to plant developmental changes [for example, cell divisions and cell enlargement such as citrus canker and hypertrophy (*4*)], thus contributing to disease symptoms. Although a number of plant targets, including susceptibility genes, have been identified (*8–10, 12–14*), the targets of most TAL effectors have not been elucidated.
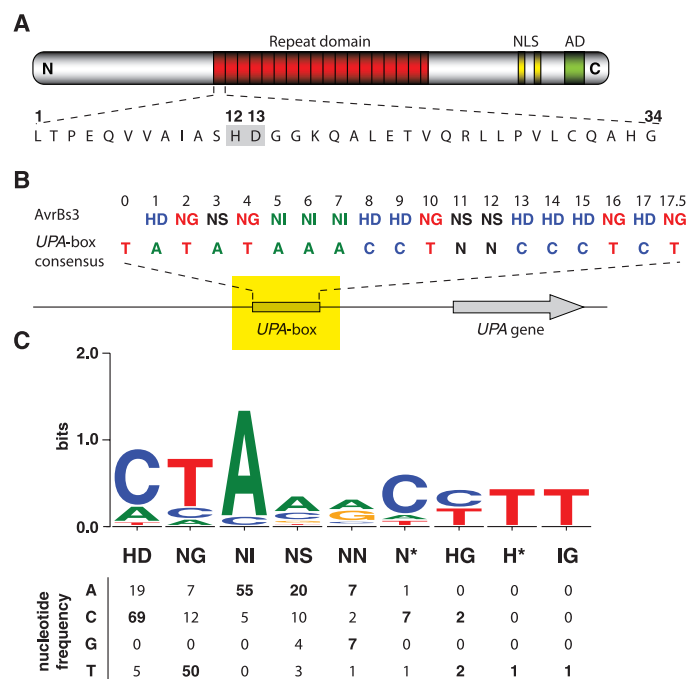
TAL effectors are characterized by a central domain of tandem repeats, nuclear localization signals (NLSs), and an acidic transcriptional activation domain (AD) (Fig. 1A) (*11, 15, 16*). Members of this effector family are highly conserved and differ mainly in the amino acid sequence and number of repeats. The number and order of repeats in a TAL effector determine its specific activity (*17, 18*). The type member of this effector family, AvrBs3 from *Xanthomonas campestris* pv. *vesicatoria*, contains 17.5 repeats and induces expression of *UPA* (upregulated by AvrBs3) genes, including the *Bs3* resistance gene in pepper plants (*9, 10, 14, 19*). The repeats of AvrBs3 are essential for DNA binding of AvrBs3 and represent a distinct type of DNA binding domain (*9*). How this domain contacts DNA and what determines specificity has remained enigmatic so far.

**A model for sequence specificity.** The fact that AvrBs3 directly binds to the *UPA* box, a promoter element in induced target genes (*9, 10*), prompted us to investigate the basis for DNA-sequence specificity. The repeat region of AvrBs3 consists of 34 amino acid repeat units that are nearly identical; however, amino acids 12 and 13 are hypervariable (Fig. 1A) (*11*). The most C-terminal repeat of AvrBs3 shows a sequence similarity to other repeats only in its first 20 amino acids and is therefore referred to as a half repeat. The repeats can be classified into different repeat types on the basis of their hypervariable 12th and 13th amino acids (Fig. 1B). Because the size of the *UPA* box [18 base pairs (bp) (*20*)

Department of Genetics, Institute of Biology, Martin-Luther-University Halle-Wittenberg, Weinbergweg 10, D-06099 Halle (Saale) Germany.

*To whom correspondence should be addressed. E-mail: jens.boch@genetik.uni-halle.de
†Present address: Sainsbury Laboratory, John Innes Centre, Norwich, Norfolk NR4 7UH, UK.
‡Present address: Icon Genetics GmbH, Biozentrum, Weinbergweg 22, D-06120 Halle (Saale), Germany.

**Fig. 1.** Model for DNA-target specificity of TAL effectors. (**A**) TAL effectors contain central tandem repeats, NLSs, and an AD. Shown is the amino acid sequence of the first repeat of AvrBs3. Hypervariable amino acids 12 and 13 are shaded in gray. (**B**) Hypervariable amino acids at position 12 and 13 of the 17.5 AvrBs3 repeats are aligned to the *UPA* box consensus (*14*). (**C**) Repeats of TAL effectors and predicted target sequences in promoters of induced genes were aligned manually. Nucleotides in the upper DNA strand that correspond to the hypervariable amino acids in each repeat were counted on the basis of the following combinations of eight effectors and experimentally identified target genes: AvrBs3/*Bs3*, UPA10, UPA12, UPA14, UPA19, UPA20, UPA21, UPA23, UPA25, AvrBs3Δrep16/*Bs3-E*, AvrBs3Δrep109/*Bs3*, AvrHah1/*Bs3*, AvrXa27/*Xa27*, PthXo1/*Xa13*, PthXo6/*OsTFX1*, and PthXo7/*OsTFIIAγ1* (fig. S1). An asterisk indicates that amino acid 13 is missing in this repeat type. Highest nucleotide frequencies are in bold. Nucleotide frequencies are displayed in a logo (http://weblogo.threeplusone.com).

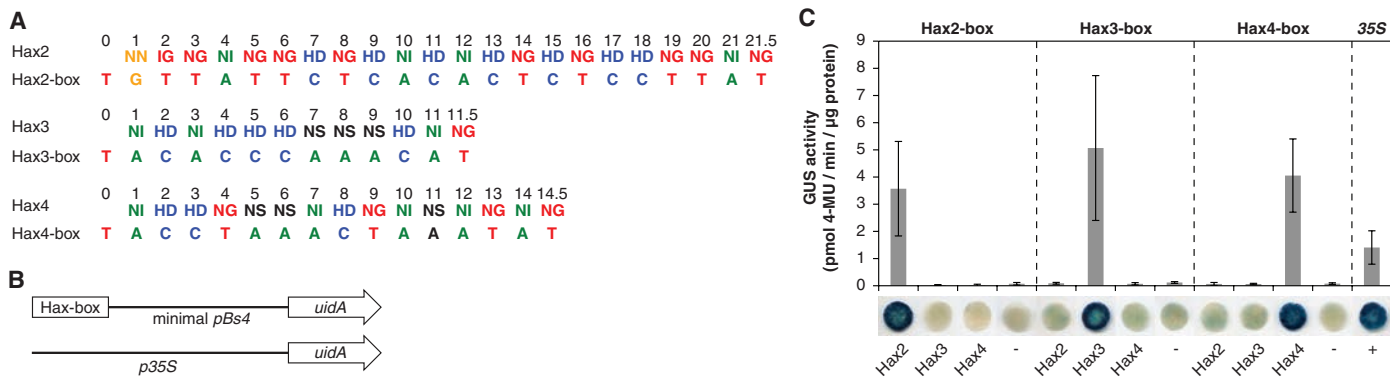| | HD | NG | NI | NS | NN | N* | HG | H* | IG |
|---|---|---|---|---|---|---|---|---|---|
| A | 19 | 7 | **55** | 20 | **7** | 1 | 0 | 0 | 0 |
| C | **69** | 12 | 5 | 10 | 2 | **7** | **2** | 0 | 0 |
| G | 0 | 0 | 0 | 4 | **7** | 0 | 0 | 0 | 0 |
| T | 5 | **50** | 0 | 3 | 1 | 1 | **2** | 1 | 1 |

**Fig. 2.** Target DNA sequences of Hax2, Hax3, and Hax4. (**A**) Amino acids 12 and 13 of the Hax2, Hax3, and Hax4 repeats and predicted target DNA specificities (Hax box). (**B**) Hax boxes were cloned in front of the minimal *Bs4* promoter into a GUS reporter vector. (**C**) Specific inducibility of the Hax boxes by Hax effectors. GUS reporter constructs were codelivered via *A. tumefaciens* into *N. benthamiana* with 35S-driven *hax2*, *hax3*, *hax4*, and empty T-DNA (−), respectively (error bars indicate SD; *n* = 3 samples). 4-MU, 4-methyl-umbelliferone. *35S::uidA* (+) served as control. Leaf discs were stained with X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide).



**Fig. 3.** DNA base pair recognition specificities of repeat types. (**A**) Hax4 and ArtX box derivatives were cloned in front of the minimal *Bs4* promoter into a GUS reporter vector. (**B**) Specificity of NG, HD, NI, and NS repeats. Hax4-inducibility of Hax4 box derivatives permutated in repeat type target bases (gray background). (**C**) Specificity of NN repeats. Artificial effector ArtX1 and predicted target DNA sequences. ArtX1-inducibility of ArtX1 box derivatives permutated in NN repeat target bases (gray background). (**D**) Artificial effectors ArtX2 and ArtX3 and derived DNA target sequences. (**E**) Specific inducibility of ArtX boxes by artificial effectors. [(A) to (E)] GUS reporter constructs were codelivered via *A. tumefaciens* into *N. benthamiana* with 35S-driven *hax4*, *artX1*, *artX2*, or *artX3* genes, and empty T-DNA (−), respectively. *35S::uidA* (+) served as control. Leaf discs were stained with X-Gluc. For quantitative data, see fig. S8.
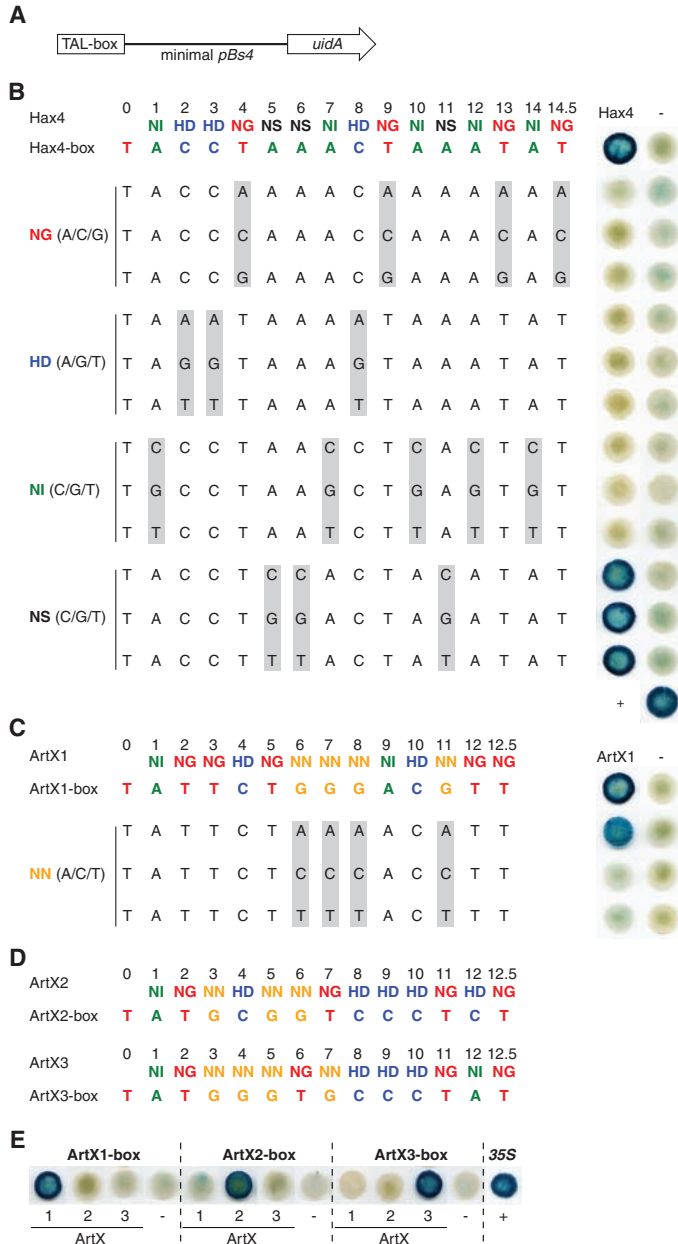
or 19 bp (*14*)] almost corresponds to the number of repeats (17.5) in AvrBs3, we considered the possibility that one repeat unit of AvrBs3 contacts one specific DNA base pair. When the repeat types of AvrBs3 (amino acid 12 and 13 of each repeat) are projected onto the *UPA* box, it becomes evident that certain repeat types correlate with specific base pairs in the target DNA. For example, HD and NI repeats have a strong preference for C and A, respectively (Fig. 1B) (*21*). For simplicity, we designate here only bases in the upper (sense) DNA strand. Our model of recognition specificity is supported by the fact that the AvrBs3 repeat deletion derivative AvrBs3Δrep16, which lacks four repeats (11 to 14) (fig. S1, A and B) (*17*), recognizes a shorter and (in the 3′ part of the box) different target DNA sequence (figs. S1 to S4) (*10, 14, 20*). On the basis of sequence comparisons of *UPA* boxes of AvrBs3-induced pepper genes (*14*) and mutational analysis (*20*), the target DNA box of AvrBs3 appeared to be 1 bp longer than the number of repeats in AvrBs3. In addition, a T is conserved at the 5′ end of the *UPA* box immediately preceding the predicted recognition specificity of the first repeat (Fig. 1). Secondary structure predictions of the AvrBs3 amino acid sequence preceding the first repeat and the repeat region (*11*) show similarities, despite the lack of sequence conservation. This suggests an additional repeat, here termed repeat 0, that extends the target boxes of AvrBs3 and possibly other TAL effectors at their 5′ ends (Fig. 1B).

**Challenging the code.** To further substantiate and extend our model (Fig. 1B), we predicted the yet unknown target DNA sequences of *Xanthomonas* TAL effectors (AvrXa27, PthXo1, PthXo6, and PthXo7 from *Xanthomonas oryzae* pv. *oryzae*) on the basis of the sequence of their repeats and inspected the promoters of known target genes and their alleles for the presence of putative binding sites. We identified sequences matching the predicted specificity in promoters of alleles that are induced in response to the corresponding TAL effector but not in noninduced alleles (fig. S1, C to
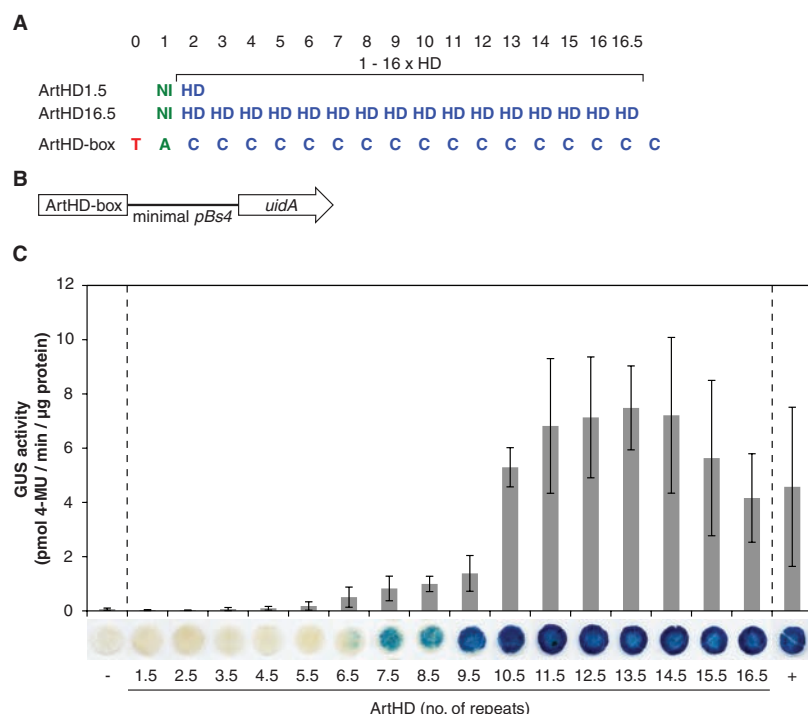
**Fig. 4.** A minimal number of repeats is required for transcriptional activation. (**A**) Artificial ArtHD effectors with different numbers (0.5 to 15.5) of HD repeats (total of 1.5 to 16.5 repeats). (**B**) An ArtHD target box consisting of TA and 17 C was cloned in front of the minimal *Bs4* promoter into a GUS reporter vector. (**C**) Promoter activation by ArtHD effectors with different number of repeats. *35S*-driven effector gene or empty T-DNA (–) were codelivered via *A. tumefaciens* with the GUS-reporter construct into *N. benthamiana* (error bars indicate SD; *n* = 3 samples 4-MU). *35S::uidA* (+) served as control. Leaf discs were stained with X-Gluc.

F). The presence of these target boxes suggests that the induced genes are direct targets of the corresponding TAL effectors. On the basis of the DNA base frequency for different repeat types in the target DNA sequences by using eight TAL effectors (fig. S1), we predicted a code for the DNA target specificity of their repeat types (Fig. 1C).

To experimentally validate our model, we predicted target DNA sequences for the TAL effectors Hax2 (21.5 repeats), Hax3 (11.5 repeats), and Hax4 (14.5 repeats) from the *Brassicaceae*-pathogen *X. campestris* pv. *armoraciae* (*22*, *23*). Hax2 is exceptional, because it has 35–amino acid instead of the typical 34–amino acid repeats. The Hax2, Hax3, and Hax4 target boxes were placed in front of the minimal (–55 to +25) tomato *Bs4* promoter, which has very weak basal activity (Fig. 2B and figs. S5 and S6) (*24*), driving a promoterless *uidA* [β-glucuronidase (*GUS*)] reporter gene. For transient expression studies, we transfected the reporter constructs together with the constitutive cauliflower mosaic virus *35S*-promoter–driven effector genes *hax2*, *hax3*, and *hax4* into *Nicotiana benthamiana* leaves using *Agrobacterium*-mediated transfer DNA (T-DNA) delivery. Qualitative and quantitative GUS assays demonstrated that promoters containing the predicted Hax2, Hax3, or Hax4 box were only induced in the presence of the corresponding effector (Fig. 2C and fig. S6). This validates our model for DNA recognition by TAL effectors and the code for DNA target specificity of their repeat

types. In addition, these data demonstrate that 35–amino acid repeats function like 34–amino acid repeats.

**Nucleotide specificity of repeat types.** Next, we addressed the importance of the first nucleotide (T; corresponding to repeat 0) in the predicted target DNA sequence for Hax3 and generated four different Hax3 boxes with either A, C, G, or T at the 5′ end (fig. S7, A and B). Codelivery of *hax3* and the reporter constructs in *N. benthamiana* demonstrated that only a promoter containing a Hax3 box with a 5′ T was induced in the presence of Hax3 (fig. S7C). This indicates that a T at position 0 contributes to promoter activation of Hax3 and probably other TAL effectors.

To address the question of whether the repeat types are specific and recognize only one specific base pair, respectively, we permutated the Hax4 box (Fig. 3, A and B). GUS assays showed that NI, HD, and NG repeats in Hax4 strongly favor recognition of A, C, and T, respectively, whereas NS repeats recognize all four base pairs (Fig. 3B and fig. S8).

Because several TAL effectors contain NN repeats (fig. S1 and table S1) (*11*) for which recognition specificity has not been tested, we generated ArtX1, an artificial TAL effector with 12.5 randomly assembled repeats that include NN repeats and deduced a corresponding DNA recognition sequence using our code (Figs. 1C and 3C). Analysis of ArtX1 box derivatives demonstrated that NN repeats recognize both A and G,

with preference for G (Fig. 3C and fig. S8). This result confirms our prediction of the Hax2 box (Fig. 2) and the natural AvrXa27 box in the promoter of the rice *Xa27* gene, which contains either an A or a G at positions corresponding to NN repeats (fig. S1C). In addition, we derived two possible AvrXa10 boxes with either A or G at positions corresponding to NN repeats in AvrXa10. Both reporter constructs were induced efficiently by AvrXa10 (fig. S9). Together, these data suggest that some repeat types favor recognition of specific base pairs, whereas others are more flexible.

**Prediction of target genes.** The expression of *hax2* in *Arabidopsis thaliana* leads to purple colored leaves, indicating an accumulation of anthocyanin (fig. S10, A and B). To identify Hax2 target genes, we analyzed promoter regions of the *A. thaliana* genome using pattern search [Patmatch, The Arabidopsis Information Resource (TAIR); www.arabidopsis.org] with degenerated Hax2 box sequences. One of the putative Hax2 target genes encodes the MYB transcription factor PAP1 (At1G56650), which controls anthocyanin biosynthesis (*25*). Pattern search revealed a suboptimal Hax2 box in the *PAP1* promoter region (fig. S10, D and E). Semiquantitative analysis of the *PAP1* transcript level demonstrated that expression of *PAP1* is strongly induced by Hax2 (fig. S10C). This demonstrates that we can successfully predict target genes for TAL effectors. On the basis of the code for repeat types (Fig. 1D) and the data described above, we predicted putative target DNA sequences for additional TAL effectors, some of which are important virulence factors (table S1).

**The number of repeats.** Because the repeat number in TAL effectors ranges from 1.5 to 28.5 (*4*), a key question is whether TAL effectors with few repeats can activate gene expression. Therefore, we tested how the number of repeats influences target gene expression. For this, we constructed N-terminal green fluorescent protein (GFP) fusions of artificial effectors that contain the N- and C-terminal regions of Hax3 and a repeat domain with 0.5 to 15.5 HD repeats (specificity for C). The HD repeats were cloned into an *Esp*3I site located after the codons for amino acid 12 and 13 of the first Hax3 repeat (NI repeat). Therefore, the first repeat in all cases was NI (specificity for A). The corresponding target DNA box consists of 17 C-residues preceded by TA (Fig. 4, A and B). Promoter activation by the artificial effectors was measured by use of the transient *Bs4*-promoter GUS assay in *N. benthamiana*. Although at least 6.5 repeats were needed for gene induction, 10.5 or more repeats led to strong reporter gene activation (Fig. 4C). Confocal microscopy revealed that all effectors localized to the plant cell nucleus, indicating production of full-length proteins. These data demonstrate that a minimal number of repeats is required to recognize the target DNA box and efficiently activate gene expression. The results also suggest that TAL effectors with fewer repeats are largely inactive.

**Artificial effectors with implications for biotechnology.** We have shown that the repeat region of TAL effectors has a sequential nature that corresponds to a consecutive target DNA sequence. Hence, it should be feasible to generate effectors with novel DNA binding specificities. Seven artificial effectors (ArtX) were generated, three with 10.5 and four with 12.5 randomly assembled repeats. They were constructed as N-terminal fusions to GFP and tested for induction of *Bs4* promoter-reporter fusions containing predicted target DNA sequences. All seven effectors induced the GUS reporter only in the presence of the corresponding target DNA box (three are shown in Fig. 3, C to E, and fig. S8). The effectors showed GFP fluorescence exclusively in the plant cell nucleus after *Agrobacterium*-mediated expression, indicating the production of full-length proteins. This shows that we are able to design DNA binding domains that target a specific DNA sequence.

Our code for recognition specificity of TAL effectors solved a 20-year enigma dating from the cloning of *avrBs3*, the first TAL effector gene (*26*). The repeat region mediates direct binding to DNA (*9*). Here, we discover how target specificity is encoded. One repeat corresponds to one base pair in the DNA, and the tandem array of repeats corresponds to a consecutive DNA sequence. Target DNA specificity is based on a two–amino acid motif per repeat, enabling the deduction of a simple code to predict the DNA target preference of TAL effectors. We have experimentally identified the following recognition preferences: HD = C; NG = T; NI = A; NS = A, C, G, or T; NN = A or G; and IG = T. Because many TAL effectors are major virulence factors (*4, 5*), the knowledge of host targets will enhance our understanding of plant disease development caused by xanthomonads.

In addition, we successfully designed artificial TAL effectors that act as transcription factors with novel DNA-binding specificities. Zinc finger transcription factors, which encode DNA binding specificity in their tandem zinc finger units, have been engineered to bind chosen DNA sequences, leading to gene control systems of great utility for biotechnology (*27, 28*). Similarly, the TAL effectors use a DNA binding code that can be exploited to generate DNA binding domains for any DNA target.

### References and Notes

1. T. Boller, S. Y. He, *Science* **324**, 742 (2009).
2. V. Göhre, S. Robatzek, *Annu. Rev. Phytopathol.* **46**, 189 (2008).
3. J. D. Jones, J. L. Dangl, *Nature* **444**, 323 (2006).
4. S. Kay, U. Bonas, *Curr. Opin. Microbiol.* **12**, 37 (2009).
5. F. F. White, B. Yang, *Plant Physiol.* **150**, 1677 (2009).
6. J. Bai, S.-H. Choi, G. Ponciano, H. Leung, J. E. Leach, *Mol. Plant Microbe Interact.* **13**, 1322 (2000).
7. B. Yang, F. F. White, *Mol. Plant Microbe Interact.* **17**, 1192 (2004).
8. K. Gu et al., *Nature* **435**, 1122 (2005).
9. S. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, *Science* **318**, 648 (2007).
10. P. Römer et al., *Science* **318**, 645 (2007).
11. S. Schornack, A. Meyer, P. Römer, T. Jordan, T. Lahaye, *J. Plant Physiol.* **163**, 256 (2006).
12. A. Sugio, B. Yang, T. Zhu, F. F. White, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10720 (2007).
13. B. Yang, A. Sugio, F. F. White, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10503 (2006).
14. S. Kay, S. Hahn, E. Marois, R. Wieduwild, U. Bonas, *Plant J.* **59**, 859 (2009).
15. G. Van den Ackerveken, E. Marois, U. Bonas, *Cell* **87**, 1307 (1996).
16. W. Zhu, B. Yang, J. M. Chittoor, L. B. Johnson, F. F. White, *Mol. Plant Microbe Interact.* **11**, 824 (1998).
17. K. Herbers, J. Conrads-Strauch, U. Bonas, *Nature* **356**, 172 (1992).
18. B. Yang, A. Sugio, F. F. White, *Mol. Plant Microbe Interact.* **18**, 142 (2005).
19. E. Marois, G. Van den Ackerveken, U. Bonas, *Mol. Plant Microbe Interact.* **15**, 637 (2002).
20. P. Römer et al., *Plant Physiol.* **150**, 1697 (2009).
21. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
22. Materials and methods are available as supporting material on *Science* Online.
23. S. Kay, J. Boch, U. Bonas, *Mol. Plant Microbe Interact.* **18**, 838 (2005).
24. S. Schornack, K. Peter, U. Bonas, T. Lahaye, *Mol. Plant Microbe Interact.* **18**, 1215 (2005).
25. J. O. Borevitz, Y. Xia, J. Blount, R. A. Dixon, C. Lamb, *Plant Cell* **12**, 2383 (2000).
26. U. Bonas, R. E. Stall, B. J. Staskawicz, *Mol. Gen. Genet.* **218**, 127 (1989).
27. J. R. Desjarlais, J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7345 (1992).
28. J. A. Townsend et al., *Nature* **459**, 442 (2009).
29. We thank J. Streubel, A. Richter, and P. Römer for kindly providing constructs; T. Gonzalez for advice on anthocyanin synthesis; and R. Kahmann for helpful suggestions on the manuscript. A patent covering the findings is pending. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SPP 1212 to J.B. and U.B. and SFB 648 to U.B.).

# REPORTS

# Modulated High-Energy Gamma-Ray Emission from the Microquasar Cygnus X-3

The Fermi LAT Collaboration*†

Microquasars are accreting black holes or neutron stars in binary systems with associated relativistic jets. Despite their frequent outburst activity, they have never been unambiguously detected emitting high-energy gamma rays. The Fermi Large Area Telescope (LAT) has detected a variable high-energy source coinciding with the position of the x-ray binary and microquasar Cygnus X-3. Its identification with Cygnus X-3 is secured by the detection of its orbital period in gamma rays, as well as the correlation of the LAT flux with radio emission from the relativistic jets of Cygnus X-3. The gamma-ray emission probably originates from within the binary system, opening new areas in which to study the formation of relativistic jets.

Cygnus X-3 (Cyg X-3) motivated to a large extent the development of high-energy gamma-ray astronomy. Detections of Cyg X-3 were reported in the 1970s and early 1980s at energies of tens of mega–electron volts, tera–electron volts, peta–electron volts, and possibly exa–electron volts, generating considerable excitement as these identified the system as a prime source of cosmic rays (*1–3*). However, the detections remained doubtful in part because observations in the late 1980s and in the 1990s by a more sensitive generation of ground-based telescopes failed to confirm the tera–electron volt and peta–electron volt detection of Cyg X-3 (*4, 5*), and also because both the COS-B satellite (*6*) and the CGRO-EGRET satellite (*7*) could not find emission from the source in the giga–electron volt range.

Here, we report the detection of a variable high-energy gamma-ray (100 MeV to 100 GeV) source that we identify with Cyg X-3 based on its location, the detection of a modulation of the gamma-ray flux at the orbital period of the binary system, and the gamma-ray variability correlated with the radio emission originating from the relativistic jets of Cyg X-3.

Cyg X-3 is a high-mass x-ray binary system (*2*) with a short orbital period of 4.8 hours (*8*), located in the Galactic plane at a distance of ~7 kpc from Earth (*9*). The nature of the compact object—neutron star or black hole—is still subject