

the dimers is different for each R-subunit isoform. In each case, however, the $\beta 4$ - $\beta 5$ loop in the R subunit appears to serve a specificity determinant for docking between the two heterodimers. By comparing the RII β tetramer with a model of the RII α holoenzyme, we see how distinct quaternary architectures can be created. It is only by comparing the full-length holoenzymes that we can appreciate that the functional nonredundancy of the PKA isoforms is also reflected in the structural nonredundancy of these isoforms. Correlating these structural differences with functional consequences will be our next challenge.

References and Notes

1. P. S. Amieux *et al.*, *J. Biol. Chem.* **277**, 27294 (2002).
2. P. S. Amieux *et al.*, *J. Biol. Chem.* **272**, 3993 (1997).
3. D. E. Cummings *et al.*, *Nature* **382**, 622 (1996).
4. S. A. Schreyer, D. E. Cummings, G. S. McKnight, R. C. LeBoeuf, *Diabetes* **50**, 2555 (2001).
5. T. A. Czyzyk, M. A. Sikorski, L. Yang, G. S. McKnight, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 276 (2008).
6. G. McConnachie, L. K. Langeberg, J. D. Scott, *Trends Mol. Med.* **12**, 317 (2006).
7. D. R. Knighton *et al.*, *Science* **253**, 407 (1991).
8. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **298**, 1912 (2002).
9. X. Cheng, C. Phelps, S. S. Taylor, *J. Biol. Chem.* **276**, 4102 (2001).
10. D. A. Johnson, P. Akamine, E. Radzio-Andzelm, M. Madhusudan, S. S. Taylor, *Chem. Rev.* **101**, 2243 (2001).
11. N. Kannan, N. Haste, S. S. Taylor, A. F. Neuwald, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1272 (2007).
12. J. M. Canaves, S. S. Taylor, *J. Mol. Evol.* **54**, 17 (2002).
13. F. S. Kinderman *et al.*, *Mol. Cell* **24**, 397 (2006).
14. M. G. Newlon *et al.*, *EMBO J.* **20**, 1651 (2001).
15. M. Diskar, H. M. Zenn, A. Kaupisch, A. Prinz, F. W. Herberg, *Cell. Signal.* **19**, 2024 (2007).
16. R. Rangel-Aldao, O. M. Rosen, *J. Biol. Chem.* **251**, 3375 (1976).
17. B. R. Martin, T. J. Deerinck, M. H. Ellisman, S. S. Taylor, R. Y. Tsien, *Chem. Biol.* **14**, 1031 (2007).
18. F. W. Herberg, S. S. Taylor, *Biochemistry* **32**, 14015 (1993).
19. S. H. Brown, J. Wu, C. Kim, K. Alberto, S. S. Taylor, *J. Mol. Biol.* **393**, 1070 (2009).
20. C. Kim, C. Y. Cheng, S. A. Saldanha, S. S. Taylor, *Cell* **130**, 1032 (2007).
21. C. Kim, N.-H. Xuong, S. S. Taylor, *Science* **307**, 690 (2005).
22. J. Wu, S. H. J. Brown, S. von Daake, S. S. Taylor, *Science* **318**, 274 (2007).
23. A. J. Boettcher *et al.*, *Structure* **19**, 265 (2011).
24. D. Vigil, D. K. Blumenthal, S. S. Taylor, J. Trehwella, *J. Mol. Biol.* **357**, 880 (2006).
25. W. T. Heller *et al.*, *J. Biol. Chem.* **279**, 19084 (2004).
26. H. M. Berman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 45 (2005).
27. T. C. Diller, N. H. Xuong, S. S. Taylor, *Protein Expr. Purif.* **20**, 357 (2000).
28. M. Batkin, I. Schwartz, S. Shaltiel, *Biochemistry* **39**, 5366 (2000).
29. J. Yang *et al.*, *J. Biol. Chem.* **284**, 6241 (2009).
30. F. W. Herberg, S. S. Taylor, W. R. Dostmann, *Biochemistry* **35**, 2934 (1996).
31. G. G. Cadd, M. D. Uhler, G. S. McKnight, *J. Biol. Chem.* **265**, 19502 (1990).
32. K. M. Zawadzki, S. S. Taylor, *J. Biol. Chem.* **279**, 7029 (2004).
33. P. Madhusudan, N. H. Akamine, S. S. Xuong, Taylor, *Nat. Struct. Biol.* **9**, 273 (2002).
34. S. F. Oliveria, M. L. Dell'Acqua, W. A. Sather, *Neuron* **55**, 261 (2007).
35. S. Manni, J. H. Mauban, C. W. Ward, M. Bond, *J. Biol. Chem.* **283**, 24145 (2008).
36. D. D. Hall *et al.*, *Biochemistry* **46**, 1635 (2007).
37. J. Zheng *et al.*, *Protein Sci.* **2**, 1559 (1993).
38. M. Gangal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12394 (1999).

Acknowledgments: We thank members of the Taylor laboratory for helpful discussions and the Advanced Light Source (beam line 8.2.2) staff of Lawrence Berkeley National Laboratory for beam access and help with data collection. This work was supported by NIH grant GM34921 (S.S.T.). Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession numbers 3TNP and 3TNQ.

Supporting Online Material

www.sciencemag.org/cgi/content/full/335/6069/712/DC1

Materials and Methods

Figs. S1 to S7

Tables S1 and S2

References (39–42)

14 September 2011; accepted 3 January 2012
10.1126/science.1213979

The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target

Amanda Nga-Sze Mak,^{1*} Philip Bradley,^{2*} Raul A. Cernadas,³
Adam J. Bogdanove,³ Barry L. Stoddard^{1†}

DNA recognition by TAL effectors is mediated by tandem repeats, each 33 to 35 residues in length, that specify nucleotides via unique repeat-variable diresidues (RVDs). The crystal structure of PthXo1 bound to its DNA target was determined by high-throughput computational structure prediction and validated by heavy-atom derivatization. Each repeat forms a left-handed, two-helix bundle that presents an RVD-containing loop to the DNA. The repeats self-associate to form a right-handed superhelix wrapped around the DNA major groove. The first RVD residue forms a stabilizing contact with the protein backbone, while the second makes a base-specific contact to the DNA sense strand. Two degenerate amino-terminal repeats also interact with the DNA. Containing several RVDs and noncanonical associations, the structure illustrates the basis of TAL effector–DNA recognition.

TAL effectors are proteins that are injected into plant cells by pathogens in the bacterial genus *Xanthomonas*. There they enter the nucleus, bind to effector-specific promoter sequences, and activate the expression of individual plant genes, which can either benefit the bacterium or trigger host defenses (1, 2). In each TAL effector, a variable number of tandem amino acid repeats (which are usually 34 residues in length), terminated by a truncated “half repeat,”

mediates DNA recognition. Each of the repeats preferentially associates with one of the four nucleotides in the target site (3, 4). The repeats are located centrally in the protein between N-terminal sequences required for bacterial type III secretion and C-terminal sequences required for nuclear localization and activation of transcription (Fig. 1A).

The nucleotide specificity of individual TAL effector repeats is encoded by two adjacent res-

idues (located at positions 12 and 13) called the repeat-variable diresidue (RVD) (Fig. 1, B and C) (4). More than 20 unique RVD sequences have been observed in TAL effectors, but just seven—HD, NG, NI, NN, NS, “N*,” (which corresponds to a 33-residue repeat in which the RVD appears to be missing its second residue), and HG—account for nearly 90% of all repeats (5) and, respectively, specify C, T, A, G/A, A/C/T/G, C/T, and T (3, 4). These relationships enable prediction of targets for existing TAL effectors and engineering of artificial TAL effectors that bind DNA sequences of choice. Consequently, TAL effectors have received much attention as DNA-targeting tools (6).

Nearly all TAL effector binding sites observed in nature are preceded by a T (3, 4). Notably, the protein sequence immediately preceding the canonical TAL effector repeats bears some similarity to the repeat consensus. It has therefore been suggested that this region of the protein may participate in DNA binding by forming a cryptic repeat structure that specifies the T (7).

¹Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, A3-025 Seattle, WA 98019, USA. ²Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M1-B514 Seattle, WA 98109, USA. ³Department of Plant Pathology and Microbiology, Iowa State University, 351 Bessey Hall, Ames, IA 50011, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: bstoddard@fhcr.org

A recent nuclear magnetic resonance structural study of 1.5 repeats of TAL effector PthA, and an accompanying small-angle x-ray scattering study of the entire protein, indicated that an isolated TAL effector repeat is largely α -helical, similar to a tetratricopeptide (TPR) fold, and that the full-length protein compacts upon DNA binding (8). However, in that study, it was unclear to what extent the structure of repeats in the context of the entire protein might differ from that of an isolated repeat, and the manner in which individual repeats associate with contiguous DNA base pairs was not resolved.

A protein construct corresponding to residues 127 to 1149 of the 23.5 repeat TAL effector PthXo1 from the rice pathogen *Xanthomonas oryzae* (Fig. 1 and fig. S1) was crystallized bound to a 36-base pair DNA duplex (table S1) containing the target sequence found in the rice genome along with flanking sequences ending in short 3' overhangs. The structure was determined with a high-throughput computational approach in which structural models built with the Rosetta software package (9) were iteratively refined and selected, guided by molecular replacement searches (fig. S2). The best model was subsequently validated with a variety of model-free features of electron density, including anomalous difference peaks calculated from a selenomethionyl derivative (fig. S3). The final structure was refined to 3.0 Å resolution to values for $R_{\text{work}}/R_{\text{free}}$ of 0.264/0.294 and excellent geometry (Table 1).

The structure consists of a relatively unperturbed B-form DNA duplex, with 23 consecutive bases of the target site intimately engaged in the major groove by a superhelical arrangement of TAL effector repeats (Fig. 2). The overall dimensions of the protein-DNA complex are approximately 60 Å by 60 Å by 90 Å. The quality of the electron density is excellent from repeat 1 through the middle of repeat 22, and then becomes less well defined.

All of the repeats in the DNA-bound PthXo1 structure form highly similar two-helix bundles (Fig. 1C). The helices span positions 3 to 11 and 14 to 33, locating the RVD in a loop between them. A proline located at position 27 creates a kink in the second helix that appears to be critical for the sequential packing and association of tandem repeats with the DNA double helix. The packing of consecutive helices within and between individual repeats is left-handed, in contrast to the right-handed packing of helices found in TPR proteins (10). The modular architecture of the TAL effector repeats is reminiscent of the mitochondrial transcription terminator mTERF (11) and the RNA-binding attenuation protein TRAP (12); however, interactions of those proteins with their nucleic acid targets are structurally distinct from those of TAL effectors with DNA and lack modular correspondence to single nucleotides.

Sequence-specific contacts of PthXo1 to the DNA are made exclusively by the second residue

in each RVD to the corresponding base on the sense strand. In contrast, the side chain at the first position of each RVD contacts the backbone carbonyl oxygen of position 8 in each repeat, constraining the RVD-containing loop (Fig. 3). Additional, nonspecific contacts to the DNA are made by a lysine and glutamine found at positions 16 and 17. The average root mean square deviation between backbone atoms in any two repeats in the PthXo1 structure is ~0.8 Å for all atoms; it is slightly greater for the 33-residue N* repeats, which are missing one residue in the RVD loops (fig. S4). The positions within the core of individual repeats are occupied entirely by small aliphatic residues, whereas several positions in the interface between repeats correspond to polar residue pairs.

The PthXo1-DNA structure displays five HD-containing repeats (all aligned to cytosines), four NG repeats and one HG repeat (aligned to thymines), one additional NG repeat aligned to cytosine, seven NI repeats (aligned with four adenines and three cytosines), two NN repeats (both opposite a guanosine), and two N* repeats paired to cytosines (Fig. 1A). The observed contacts by individual repeats (Fig. 3) correlate well with their specificity and fidelity (or lack thereof) that have been described via bioinformatic and genetic analyses. The sole NS in PthXo1 and one additional N* are located in the last full repeat and the half repeat, respectively, which are disordered in the structure.

In the HD RVDs, the aspartate residue makes van der Waals contacts with the edge of the corresponding cytosine base and a hydrogen bond to the cytosine N4 atom. Contacts between cytosine bases in protein-DNA complexes and charged acidic side chains, which exclude alternative base identities via physical and electrostatic clash, have been observed in a wide variety of solved sequence-specific protein-DNA complexes (13).

Both the NG and HG repeats make a contact in which the backbone α carbon of the glycine residue forms a nonpolar van der Waals interaction with the methyl group of the opposing thymine base (average distance ~3.3 Å). At the one position where an NG is aligned opposite a cytosine base, the backbone carbonyl and α carbon of the same glycine residue displays a less favorable, far more distant contact (~6 Å).

The second asparagine residue in the NN RVDs is positioned to make a hydrogen bond with the N7 nitrogen of an opposing guanine base. This RVD associates with either guanosine or adenine with roughly equal frequency (3, 4, 14); the availability of an N7 nitrogen in either purine ring appears to explain that observation (13).

PthXo1 contains two 33-residue N* repeats (7 and 22). Because RVDs are followed immediately by two conserved glycine residues, this repeat is equivalent to an NG repeat in which one of those glycine residues is missing. The crystal structure indicates that the deletion results

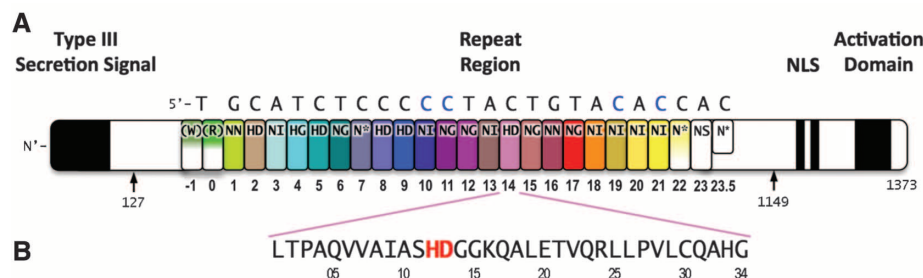


Fig. 1. (A and B) Domain organization of PthXo1 and (C) structure of a single TAL effector repeat. TAL effectors contain N-terminal signals for bacterial type III secretion, tandem repeats that specify the target nucleotide sequence, nuclear localization signals (NLS), and a C-terminal region that is required for transcriptional activation. PthXo1 contains 23.5 canonical repeats (color coded to match Fig. 2) that contact the DNA target found in the promoter of the rice *Os8N3* gene (15). 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 (3, 4). Arrows indicate the start and end of the crystallized protein construct. In the structure, repeats 22 to 23.5 are poorly ordered, as are the C termini of the two N-terminal cryptic repeats. The sequence and structure of a representative repeat (number 14) is shown; RVD residues (HD) that recognize cytosine are shown in red. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; G, Gly; H, His; I, Ile; K, Lys; L, Leu; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; and V, Val.

in a truncated RVD loop that extends less deeply into the DNA major groove, with the glycine at position 13 located a considerable distance (>6 Å) from the corresponding sense strand base. Consistent with this observation, the observed specificity of the N* repeat is relatively lax (4).

Finally, NI, which is the second most common RVD overall, accounting for roughly 20% of all TAL effector repeats, occurs seven times in PthXo1 and displays an unusual contact pattern to adenosine or cytosine bases. The aliphatic side chain of the isoleucine residue is observed to make nonpolar van der Waals contacts to C8 (and N7) of the adenine purine ring, or to C5 of the cytosine pyrimidine ring. These contacts would appear to necessitate desolvation of at least one polar atom in the adenosine ring, without the formation of a compensating hydrogen bond, and might therefore reasonably be expected to represent a reduced-affinity interaction.

N-terminal to the canonical repeats, the PthXo1 structure reveals two degenerate repeat folds that appear to cooperate to specify the conserved thymine that precedes the RVD-specified sequence (Fig. 4). We have designated these as the 0th and –1st repeats. Residues 221 to 239 and residues 256 to 273 each form a helix and an adjoining loop that resembles helix 1 and the RVD loop in the canonical repeats; the remaining residues in each region are poorly ordered. Those two N-terminal regions converge near the 5' thymine base, with the indole ring of tryptophan 232 (in the –1st repeat) making a van der Waals contact with the methyl group of that base. Mutation of the thymine reduces TAL effector activity at the target (3, 15). Tryptophan-232, as well as the surrounding residues, is highly conserved across available, intact TAL effector sequences. Some TAL effectors efficiently target sequences preceded by a cytosine rather than a thymine (14, 16). Though less favorable, the packing of tryptophan 232 would be expected to accommodate this substitution.

In addition to revealing folding and interactions of the N-terminal cryptic repeats with the 5' end of the DNA target site and illustrating the functions of the six most common repeat types in TAL effector–DNA recognition, the structure provides a basis for prediction of structures that are not represented. For example, an alignment of the 35-residue repeat type found in some TAL effectors with the more common 34-residue repeat type found in PthXo1 (fig. S5) indicates that the additional residue (a proline) at position 33 would be located within the relatively disordered turn region that connects the helices of one repeat to the next. The 35-residue repeat therefore can be predicted to be functionally indistinguishable from the 34. Likewise, although the sole NS repeat in PthXo1 is in an apparently disordered part of the protein–DNA complex, the overall homogeneity of the repeat structures and the consistent role of the first RVD residue in stabilizing the

Table 1. Crystallographic data and refinement statistics. WT, wild type; SeMet, selenomethionine; RMSD, root mean square deviation; ALS, Advanced Light Source; APS, Advanced Photon Source.

Data statistics		
Dataset	WT	SeMet
X-ray source	ALS 5.0.2	APS 21-ID-F
Wavelength (Å)	1.177	1.378
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	$a = 95.6$ $b = 248.5$ $c = 54.6$	$a = 100.7$ $b = 247.8$ $c = 54.2$
Resolution (Å)	50–3.0 (3.11–3.0)	50–4.0 (4.14–4.0)
R_{merge} (%)	0.121 (0.431)	0.087 (0.139)
$I/\sigma(I)$	9.3 (3.5)	10.3 (3.7)
Redundancy	5.6 (4.9)	4.9 (5.0)
Completeness (%)	96.6 (90.4)	95.8 (97.7)
Mosaicity (°)	0.8	0.8
Unique reflections	25841	11591
Refinement		
R_{work}	0.264	
R_{free}	0.294	
Protein atoms	6086	
DNA atoms	1552	
Heteroatoms (waters)	216	
RMSD bond lengths (Å)	0.021	
RMSD bond angles (°)	2.4	
Average B factor (Å ²)	85.1	
Ramachandran	73.6%, 26.4%, (% core, allowed, generous, disallowed)	0%, 0%

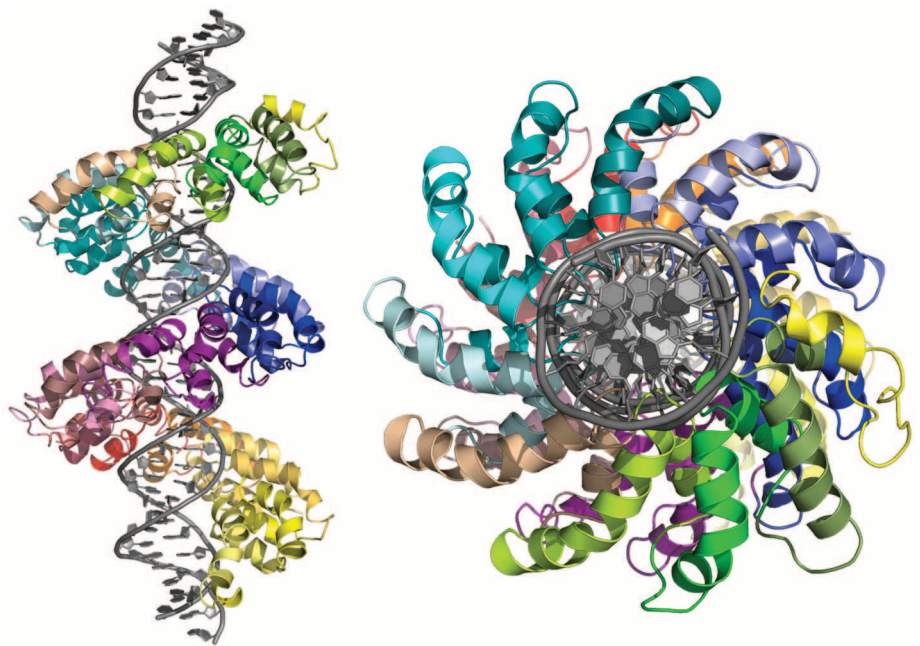


Fig. 2. Structure of the PthXo1 DNA binding region in complex with its target site. The coloring of individual repeats matches the schematic in Fig. 1.

RVD loop to facilitate specific contacts of the second residue with the DNA should make it possible to computationally model the potential nucleotide interactions of NS, as well as those of rare or artificial RVDs.

The protein–DNA complex studied leaves some questions unanswered, such as the structure of the N- and C-terminal portions of TAL effectors that are, respectively, required for translocation and interaction with host transcriptional

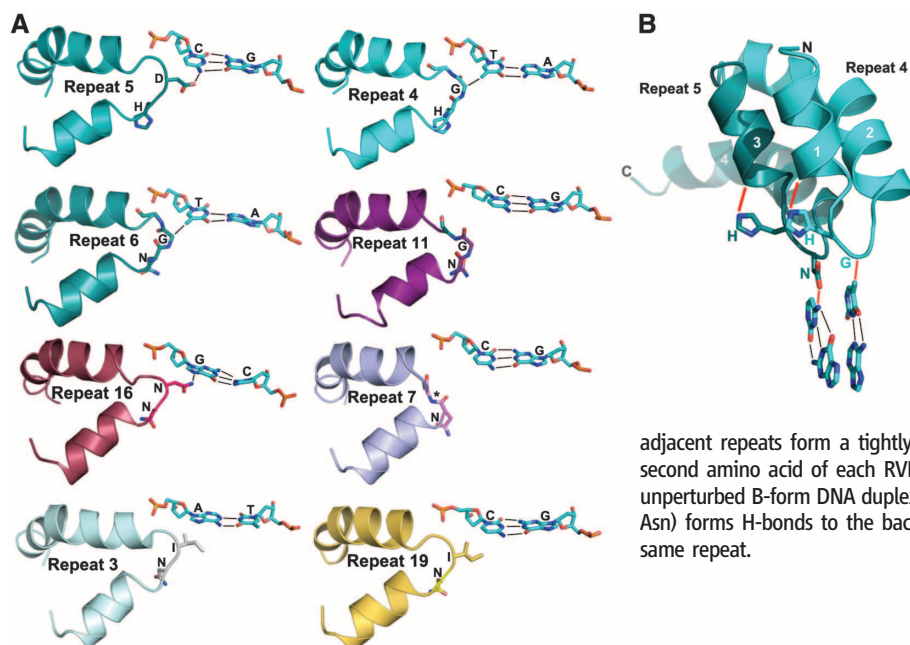


Fig. 3. Topology and contacts between TAL effector repeats and DNA bases. **(A)** Eight distinct combinations of RVDs and DNA bases are observed in the structure. HD (repeat 5) forms a steric and electrostatic contact with cytosine; HG (repeat 4) and NG (repeat 6) both form nonpolar interactions between the glycine α -carbon and the thymine methyl group. A “mismatch” between NG and a cytosine (repeat 11) results in a longer distance from the RVD to the base. NN associates with either guanine (repeat 16) or with adenine (which would interact with the same N7 nitrogen of the purine base). NI forms a desolvating interface with either adenine (repeat 3) or cytosine (repeat 19). The reduction in loop length by one residue in the N* RVD (repeat 7) results in an increased distance to the base. **(B)** Two

adjacent repeats form a tightly packed left-handed bundle of helices that position the second amino acid of each RVD in proximity to corresponding consecutive bases in an unperturbed B-form DNA duplex. The first residue of each RVD (position 12, either His or Asn) forms H-bonds to the backbone carbonyl oxygen of amino acid position 8 of the same repeat.

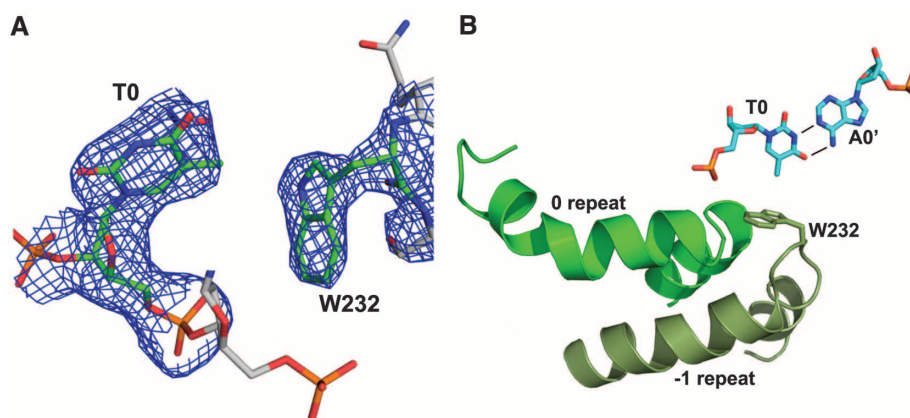


Fig. 4. N-terminal cryptic repeats and contacts with 5' thymine. **(A)** 2Fo-Fc electron density maps contoured around thymine at position “0” and tryptophan 232 in the “-1” repeat. **(B)** Residues 221 to 239 and residues 256 to 273 each form a helix and an adjoining loop that resembles helix 1 and the RVD loop in the canonical repeats; the remaining residues in each region are poorly ordered. W232 forms a nonpolar van der Waals contact with the methyl carbon of the thymine base at position 0.

machinery. As well, because of the observed disorder at either end, it does not yet precisely define the minimal TAL effector–DNA binding domain. However, by demonstrating the essential features that accomplish interaction specificity, the structure provides a foundation for more accurately predicting and efficiently exploiting TAL effector–DNA targeting. More fundamentally, it reveals the hitherto enigmatic structural nature of a simple solution that an important group of pathogens has evolved to manipulate host gene expression in a specific yet highly adaptable manner.

References and Notes

1. S. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, *Science* **318**, 648 (2007).
2. P. Römer *et al.*, *Science* **318**, 645 (2007).
3. J. Boch *et al.*, *Science* **326**, 1509 (2009).
4. M. J. Moscou, A. J. Bogdanove, *Science* **326**, 1501 (2009).
5. J. Boch, U. Bonas, *Annu. Rev. Phytopathol.* **48**, 419 (2010).
6. A. J. Bogdanove, D. F. Voytas, *Science* **333**, 1843 (2011).
7. A. J. Bogdanove, S. Schornack, T. Lahaye, *Curr. Opin. Plant Biol.* **13**, 394 (2010).
8. M. T. Murakami *et al.*, *Proteins Struct. Funct. Bioinf.* **78**, 3386 (2010).
9. A. Leaver-Fay *et al.*, *Methods Enzymol.* **487**, 545 (2011).
10. L. D. D’Andrea, L. Regan, *Trends Biochem. Sci.* **28**, 655 (2003).
11. N. Jiménez-Menéndez *et al.*, *Nat. Struct. Mol. Biol.* **17**, 891 (2010).
12. N. H. Hopcroft *et al.*, *J. Mol. Biol.* **338**, 43 (2004).
13. R. Rohs *et al.*, *Annu. Rev. Biochem.* **79**, 233 (2010).
14. J. C. Miller *et al.*, *Nat. Biotechnol.* **29**, 143 (2011).
15. P. Römer *et al.*, *New Phytol.* **187**, 1048 (2010).
16. Y. Yu *et al.*, *Mol. Plant Microbe Interact.* **24**, 1102 (2011).

Acknowledgments: This project was funded by NIH (grants RL1 OC833133 to B.L.S., R01GM098861 to B.L.S. and A.J.B., and R01 GM088277 to P.H.B.), NSF grant 0820831 to A.J.B.,

and a Seales Scholars Fellowship to P.H.B. A.N.S.-M. was supported by a training grant from the Northwest Genome Engineering Consortium. We thank the staff of the Advanced Light Source beamline 5.0.2 and L. Doyle, B. Shen, R. Takeuchi, J. Bolduc, and C. Schmidt for technical assistance and advice; T. Edwards and M. Clifton for collecting SeMet data; and C. Pabo for helpful discussion. A.J.B. is an inventor on a patent application titled “TAL effector-mediated DNA modification” (US-2011/0145940-A1 and PCT/US2010/059932). This intellectual property, co-owned by Iowa State University and the University of Minnesota, has been licensed to Collectis. The refined coordinates and corresponding x-ray intensities for the PthXo1–DNA structure have been deposited in the RCSB Protein Database (accession code 3UGM).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1216211/DC1
Materials and Methods
Figs. S1 to S5
Table S1
References (17–29)

4 November 2011; accepted 21 December 2011
Published online 5 January 2012;
10.1126/science.1216211