**Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors**

Dong Deng,1,2,*, Chuangye Yan,2,*, Xiaojing Pan,1,2,† Magdy Mahfouz,3 Jiawei Wang,1 Jian-Kang Zhu,1 Yigong Shi,1,2† Nieng Yan1,2†

**Abstract**

TAL (transcription activator–like) effectors, secreted by phytopathogenic bacteria, recognize host DNA sequences through a central domain of tandem repeats. Each repeat comprises 33 to 35 conserved amino acids and targets a specific base pair by using two hypervariable residues [known as repeat variable diresidues (RVDs)] at positions 12 and 13. Here, we report the crystal structures of an 11.5-repeat TAL effector in both DNA-free and DNA-bound states. Each TAL repeat comprises two helices connected by a short RVD-containing loop. The 11.5 repeats form a right-handed, superhelical structure that tracks along the sense strand of DNA duplex, with RVDs contacting the major groove. The 12th residue stabilizes the RVD loop, whereas the 13th residue makes a base-specific contact. Understanding DNA recognition by TAL effectors may facilitate rational design of DNA-binding proteins with biotechnological applications.

1State Key Laboratory of Bio-Membrane and Membrane Biotechnology, Tsinghua University, Beijing 100084, China. 2Tsinghua-Peking Center for Life Sciences, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. 3Center for Plant Stress Genomics and Technology, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia. 4Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: shi-lab@tsinghua.edu.cn (Y.S.; nyan@tsinghua.edu.cn (N.Y.)

**Fig. 1.** Structure of the TAL repeats in DNA-free dHax3. (A) The 11 TAL repeats of dHax3 form a right-handed superhelical assembly. Two perpendicular views are presented with the RVDs highlighted in red in the right image. (B) All TAL repeats exhibit a nearly identical conformation. Each repeat is organized into short (a) and long (b) α helices connected by a short loop where the two RVDs at positions 12 and 13 are located. All structure figures were prepared with PyMOL (30). 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.
α-helical repeat modules such as HEAT (22) and TPR (23), in which the two helices are nearly parallel to each other. A nuclear magnetic resonance (NMR) structure of 1.5 TAL repeats in the protein PthA was previously reported (24); however, our TAL repeat structure exhibits major differences from that in PthA (figs. S2, D and E).

The 11 TAL repeats of dHax3 complete a full helical turn; the RVD loops form the innermost spiral with a pitch of 60 Å per turn (Fig. 1A). The 11 α helices form an internal layer along the superhelical axis, whereas the 11 β helices constitute an external layer (Fig. 1A). These structural features suggest a DNA-binding model in which the DNA molecule is placed within the TAL superhelical assembly along the axis.

We crystallized a binary complex between dHax3 (residues 231 to 720), which encompasses the entire 11.5 TAL repeats, and a 17-base pair (bp) DNA binding element (20), with 5′-TGTCCTATTATCTCTCT-3′ as the sense strand. The structure was determined by molecular replacement at 1.85 Å resolution (table S2 and fig. S3A). There are two complexes in each asymmetric unit (fig. S3B). The two protein molecules (designated A and B) can be superimposed with a root mean square deviation (RMSD) of 1.2 Å over 447 Ca atoms (fig. S3C). Because these two complexes exhibit identical features for most repeats, we mainly describe structural analysis on molecule A.

In the complex structure, dHax3 comprises 12 repeats (residues 289 to 691), with the C-terminal 0.5 repeat contributed by nonconserved amino acids (Fig. 2). These repeats are capped by three and two short α helices at the N and C termini, respectively (Fig. 2). Similar to DNA-free dHax3, all repeats exhibit a nearly identical conformation except RVD loops in repeat 6 of molecule A and repeat 5 of molecule B (figs. S3D and S4). The superhelical dHax3 structure tracks along the major groove of the DNA duplex. The conformation of the 17-bp DNA is largely B-form (table S3), with 11 base pairs per turn and a pitch of about 35 Å.

In the structures of both DNA-free and DNA-bound dHax3, there are 11 TAL repeats per helical turn (Fig. 3A). Comparison of any corresponding repeat between these two structures reveals little difference, with an RMSD of 0.25 to 0.34 Å over about 30 Ca atoms (Fig. 3B). However, the superhelical pitch is reduced from 60 Å in DNA-free form to about 35 Å in the DNA-bound form (Fig. 3A). Whereas the main chains of the first 22 amino acids are precisely superimposed, subtle conformational variations accumulate for residues 23 to 34, resulting in notable differences between the positions of the Ca atoms in Gly24 (Fig. 3B). Such differences are gradually amplified over an increasing number of repeats (fig. S5), ultimately resulting in the compression of the superhelical assembly in the DNA-bound form. Such conformational plasticity is consistent with the predominantly van der Waals interactions between adjacent TAL repeats, which can tolerate minor distance shifts (Fig. 3, C and D, and fig. S6). The ability to undergo substantial conformational changes appears to be a conserved feature for superhelical assemblies exemplified by Armadillo repeats in β-catenin (25) and HEAT repeats in keryopherin α (26) and the scaffold subunit of protein phosphatase 2A (PP2A) (27). The conformational plasticity of the TAL repeats, which was previously noted (24), is likely essential for the function of TALEs.

Analysis of the electrostatic surface potential reveals a stripe of positively charged amino acids along the inner ridge of the dHax3 superhelical assembly (Fig. 4A and fig. S7A). Each phosphate group in the sense strand of the DNA duplex is accommodated in a shallow surface pocket along the basic stripe (Fig. 4A, left). Lys16 and Glu17, which are located at the beginning of helix b in each repeat, contribute to the positive electrostatic potential for interaction with the negatively charged phosphate (Fig. 4A, right). Interaction with the phosphate group of DNA duplex, invariant among repeats 1 through 11, is mediated by hydrogen bonds (fig. S7, B and C).

The two hypervariable residues in the RVD loops, positioned in close proximity to the sense strand in the DNA major groove (Fig. 4B), play different biochemical roles. Residue 12, either His or Asn in the 11.5 TAL repeats of dHax3 (fig. S1), does not directly contact DNA. Instead, the side chains of His12 and Asn12 point away from DNA and located at the C-terminal end of helix a in each TAL repeat (Fig. 4C). Thus, the primary role of residue 12 in TAL repeats is not to directly recognize DNA but to stabilize the local conformation to adjacent repeats.
of the RVD loops. Supporting this analysis, there is a water-mediated H bond between the imidazole group of His12 in TAL repeat 1 and the carboxylate oxygen atom of Asp13 in repeat 2 (Fig. 4C). Identical interaction is observed between His12 of repeat 2 and Asp13 of repeat 3. These structural findings demonstrate that His12 or Asn12 contributes indirectly to DNA binding by stabilizing the proper conformation of the RVD loops, which allows residue 13 to specifically recognize DNA bases.

Among the more than 20 codes identified for DNA recognition by TALE RVDs, some are more frequently observed than others (1, 18). The TAL repeats in dHax3 use three codes, in which the two hypervariable residues HD, NG, and NS specifically recognize the DNA bases C, T, and A, respectively (20). These three codes account for about half of all cases reported (1). The structure of DNA-bound dHax3 provides a satisfying explanation to these codes. In the case of HD→C, the carboxylate oxygen atom of Asp13 accepts a H bond from the amine group of cytosine in TAL repeats 1 to 3, 9, and 11 (Fig. 4D). In the case of NS→A, the hydroxyl group of Ser13 in TAL repeat 7 donates a H bond to the N7 atom of adenine (Fig. 4D). Compared with HD, NS is nonselective in that it can recognize all four bases (14). Similar to adenine, guanine also contains a N7 atom, which is likely recognized by Ser13 in the same manner. Recognition of cytosine or thymine may require a slightly different conformation of the RVD loop, a scenario awaiting further structural evidence.

The correlation between NG and the base T is intriguing. Instead of providing any specific interaction, the placement of Gly at position 13 allows sufficient space to accommodate the 5-methyl group of thymine (Fig. 4E). In TAL repeats 4, 8, 10, and 12, the distance between the Ca of Gly13 and the 5-methyl group of thymine is between 3.4 and 3.7 Å, allowing van der Waals interaction. Substitution of Gly with any other residue would likely introduce steric clash with the 5-methyl group of thymine, providing a structural explanation for the observation that recognition of the base T usually requires Gly at position 13 (1). However, in repeats 5 of molecule A and 6 of molecule B, the distance between Gly-Ca and the 5-methyl group of thymine is more than 5 Å. We speculate that mutation of Gly13 to an amino acid with a short side chain may be tolerated here.

Both the structure and the mode of DNA binding by the TAL repeats differ from those of other known DNA-binding domains such as zinc-finger domain, basic leucine zipper motif, and helix-turn-helix motif (fig. S8). The modular nature of the DNA-TAL repeats is also different from that of known RNA-binding proteins such as trp RNA-binding attenuation protein (TRAP) (28). The closest entry from an exhaustive search of the Protein Data Bank (PDB) using DALI (29) is the structure of DNA-bound MTERF1 (mitochondria transcription terminator 1) (fig. S8), which also exhibits a superhelical conformation.

Fig. 4.
and has a Z score of 7.0 and RMSD of 3.2 Å over 184 aligned Co atoms with dHax3. However, the MTERF motif comprises two α helices and one 310-helix, with considerable conformational variation among repeats. In addition, MTERF1 binding results in substantial unwinding of DNA duplex (fig. S8).

Our structural investigation provides explanation for about half of the frequently used codes for DNA recognition by TAL repeats. Among the remaining codes, how N1 and NN recognize the bases A and G/A, respectively, remains to be elucidated. We suspect that the second Asn residue of NN may favor G/A through a specific H bond. Some of the less frequently used codes can also be explained by our available structural information. For example, explanation for the code ND→C should be similar to that for HD→C, which was observed here (Fig. 4D). On the other hand, rationalization for the code XG→T is likely the same as that for NG→T (Fig. 4E). Because 5′-methylcytosine is similar to T, we suspect that XG might also be able to recognize 5′-methylcytosine.

Our study represents a step toward comprehensive rationalization of sequence-specific DNA recognition by TAL repeats. Many questions remain. It is yet to be seen whether the arrangement of 11 repeats per turn is unique to dHax3 or a common feature of all TAL repeats. Although the base T is required for repeat “O” (J4, 26), our structure of DNA-bound dHax3 does not provide an intuitive clue, because T at position zero is not particularly coordinated by either the N-terminal domain or the adjacent repeats (fig. S9). Nonetheless, visualization of the modular, base-specific recognition by the TAL repeats may greatly facilitate rational design of novel DNA-binding proteins with a range of pragmatic applications.

References and Notes
21. See supporting material on Science Online.

Acknowledgments: We thank J. He and Q. Wang at Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U and K. Hasegawa and T. Kumasaka at the Spring-8 beamline BL41XU for on-site assistance. This work was supported by funds from the Ministry of Science and Technology (grant numbers 2009CB188801, 2009CB189802, and 2011CB910901); projects 30880001, 20101001, and 31070644 of the National Natural Science Foundation of China; and Tsinghua University 985 Phase II funds. Coordinates and structure factors for the DNA-free and DNA-bound structures have been deposited with the Protein Data Bank under accession codes 3V6P and 3V6T.

Supporting Online Material
www.sciencemag.org/cgi/content/full/science.1215670/DC1 Materials and Methods
Materials and Methods
Figs. S1 to S9
Tables S1 to S3
References (31–44)
24 October 2011; accepted 21 December 2011
Published online 5 January 2012;
10.1126/science.1215670

Rescued Tolerant CD8 T Cells Are Preprogrammed to Reestablish the Tolerant State

Andrea Schietinger,1,2 Jeffrey J. Delrow,3 Ryan S. Basom,3 Joseph N. Blattman,1,2,4† Philip D. Greenberg1,4†

Tolerant self-antigen–specific CD8 T cells fail to proliferate in response to antigen, thereby preventing autoimmune disease. By using an in vivo mouse model, we show that tolerant T cells proliferate and become functional under lymphopenic conditions, even in a tolerogenic environment. However, T cell rescue is only transient, with tolerance re imposed upon lymphorepletion even in the absence of tolerogen (self-antigen), challenging the prevailing paradigm that continuous antigen exposure is critical to maintain tolerance. Genome-wide messenger RNA and microRNA profiling revealed that tolerant T cells have a tolerance-specific gene profile that can be temporarily overridden under lymphopenic conditions but is inevitably re imposed, which suggests epigenetic regulation. These insights into the regulatory mechanisms that maintain or break self-tolerance may lead to new strategies for the treatment of cancer and autoimmunity.

1Department of Immunology, University of Washington (UW), Seattle, WA 98195, USA.
2Program of Immunology, Fred Hutchinson Cancer Research Center (FHCRC), Seattle, WA 98109, USA.
3Genomics Shared Resource, FHCRC, Seattle, WA 98109, USA.
4*Present address: School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA.
†To whom correspondence should be addressed. E-mail: pgreen@u.washington.edu
Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors
Dong Deng et al.
Science 335, 720 (2012);
DOI: 10.1126/science.1215670

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

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of February 26, 2016):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
/content/335/6069/720.full.html

Supporting Online Material can be found at:
/content/suppl/2012/01/05/science.1215670.DC1.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
/content/335/6069/720.full.html#related

This article cites 41 articles, 13 of which can be accessed free:
/content/335/6069/720.full.html#ref-list-1

This article has been cited by 59 articles hosted by HighWire Press; see:
/content/335/6069/720.full.html#related-urls

This article appears in the following subject collections:
Biochemistry
/cgi/collection/biochem

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2012 by the American Association for the Advancement of Science; all rights reserved. The title Science is a registered trademark of AAAS.