4308–4318 *Nucleic Acids Research, 2019, Vol. 47, No. 8* *Published online 20 February 2019*
doi: 10.1093/nar/gkz113

Structural basis of dimerization and dual W-box DNA recognition by rice WRKY domain

Xiankun Cheng[1], Yanxiang Zhao[2], Qingshan Jiang[1], Jun Yang[1,3], Wensheng Zhao[1,3], Ian A. Taylor[4], You-Liang Peng[1,3], Dongli Wang[5,*] and Junfeng Liu[1,*]

[1]MOA Key Laboratory of Plant Pathology, joint international Research Laboratory of Crop Molecular Breeding, College of Plant Protection, China Agricultural University, Beijing 100193, China, [2]College of Plant Health and Medicine, and Key Lab of Integrated Crop Disease and Pest Management of Shandong Province, Qingdao Agricultural University, Qingdao 266109, China, [3]State key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China, [4]Macromolecular Structure Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK and [5]Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

Received November 12, 2018; Revised February 10, 2019; Editorial Decision February 11, 2019; Accepted February 14, 2019

ABSTRACT

In rice, the critical regulator of the salicylic acid signalling pathway is OsWRKY45, a transcription factor (TF) of the WRKY TF family that functions by binding to the W-box of gene promoters, but the structural basis of OsWRKY45/W-box DNA recognition is unknown. Here, we show the crystal structure of the DNA binding domain of OsWRKY45 (OsWRKY45–DBD, i.e. the WRKY and zinc finger domain) in complex with a W-box DNA. Surprisingly, two OsWRKY45–DBD molecules exchange β4-β5 strands to form a dimer. The domain swapping occurs at the hinge region between the β3 and β4 strands, and is bridged and stabilized by zinc ion via coordinating residues from different chains. The dimer contains two identical DNA binding domains that interact with the major groove of W-box DNA. In addition to hydrophobic and direct hydrogen bonds, water mediated hydrogen bonds are also involved in base-specific interaction between protein and DNA. Finally, we discussed the cause and consequence of domain swapping of OsWRKY45–DBD, and based on our work and that of previous studies present a detailed mechanism of W-box recognition by WRKY TFs. This work reveals a novel dimerization and DNA-binding mode of WRKY TFs, and an intricate picture of the WRKY/W-box DNA recognition.

INTRODUCTION

Plants have evolved sophisticated system to counteract biotic and abiotic stresses. The salicylic acid (SA) signalling pathway is used by plants for biotrophic pathogen defence—including local defence and systemic acquired resistance—and the regulation of physiological and biochemical processes during the plant life span (1). In the model plant *Arabidopsis thaliana*, the SA pathway is mainly regulated by the transcriptional coactivator NPR1 (2); but in rice (*Oryza sativa*), it has evolved into two sub-pathways, WRKY45-dependent and OsNPR1-dependent (3). Difference of the two sub-pathways is that most of the WRKY45-dependent benzothiadiazole-responsive (BTH, a coactivator of SA) genes can be upregulated by BTH (3), while more than half of the OsNPR1-dependent BTH-responsive genes can be downregulated by BTH (4). Knockdown of OsWRKY45 can severely reduce rice resistance to fungal pathogen *Magnaporthe oryzae* and bacterial pathogen *Xanthomonas oryzae*, and overexpression of OsWRKY45 results in an increased resistance to the two pathogens, indicating the pivotal role of WRKY45-dependent SA signalling pathway in rice (5).

OsWRKY45 belongs to group IIIa of the plant specified WRKY transcription factor (TF) family (6). All the WRKY TFs contain the WRKY domain, which is ∼60 residues long and contains an N-terminal WRKYGQK sequence required for W-box (TTGACC/T) DNA recognition and a C-terminal zinc finger motif for structure stabilization. There are >100 WRKY genes in rice and more than 80 in Arabidopsis, and based on sequence homology of WRKY domains, WRKY proteins are divided into groups I to III and each group is further divided into several subgroups (7,8). The zinc finger of WRKY TFs is $CX_{4-5}CX_{22-23}HXH$ (C2H2) (group Ia and II) or $CX_7CX_{23}HXC$ (C2HC) (group Ib and III). There are two OsWRKY45 allelic genes in rice, OsWRKY45-1 in the Japonica subspecies and OsWRKY45-2 in the indica sub-

*To whom correspondence should be addressed. Junfeng Liu. Tel: +86 10 62733607; Fax: +86 10 62733607; Email: jliu@cau.edu.cn
Correspondence may also be addressed to Dongli Wang. Email: wangdl@mail.tsinghua.edu.cn

species, and there is one amino acid difference (Leu150 of OsWRKY45-1 and Met150 of OsWRKY45-2) in their WRKY domains (5).

A WRKY45-dependent SA signalling cascade in rice has been proposed: in response to pathogen, SA, facilitated by BTH and probably through OsMKK10-2 (MAPK kinase), activates OsMPK6 (MAP kinase) by Thr/Tyr dual phosphorylation; the activated OsMPK6 in turn activates OsWRKY45 by phosphorylating one or two serine residues of its C-terminal; then the activated OsWRKY45 binds to the W-box DNA of target gene promoters and assists the initiation of gene transcription (9). OsWRKY45 also upregulates its own expression level (3). Useless or excess OsWRKY45 is degraded by the ubiquitin proteasome system (10). OsWRKY45 can interact with other proteins. The N-terminal of OsWRKY45 interacts with N-terminal coiled-coil (CC) domain of the Pb1 (a panicle blast resistance gene) protein in the nucleus and confers resistance to *M. oryzae* (10). OsWRKY45 can also form a heterodimer with OsWRKY62 during pathogen infection and the dimer can activate the transcription of the DPF gene (diterpenoid phytoalexin biosynthetic gene) (11). OsWRKY45 is also involved in abiotic stresses such as low temperature, since rice use the abscisic acid (ABA) signalling pathway to counteract abiotic stresses, and ABA causes OsPTP1/2 to inactive OsMPK6 by dephosphorylation, thus decrease the activity of OsWRKY45 (9).

Despite its central role in rice SA signalling regulation, no structural information of OsWRKY45 or its binding property to W-box DNA has been reported. Other published structures of WRKY proteins include the AtWRKY1-DBD monomer crystal structure (12), the AtWRKY4-DBD apo (13) and AtWRKY4-DBD/W-box DNA complex solution NMR structures (14) and the WRKY domain of RRS1 (Resistance to *Ralstonia solanacearum* 1, RRS1-$R_{WRKY,}$ gene name AtWRKY52) (15). However, all of these WRKY proteins have the C2H2 zinc finger, thus no structural information of the C2HC zinc finger and its interaction with W-box DNA available. In addition, homo- or hetero- dimer or oligomer of WRKY TFs are prevalent in plants but no structural study of a WRKY oligomer has been reported. Here, we present the crystal structure of OsWRKY45–DBD/DNA complex and reveal an unexpected zinc-bridged OsWRKY45–DBD homodimer architecture. These data combined with previous studies now provide the precise molecular details of the WRKY/W-box DNA recognition mechanism.

MATERIALS AND METHODS

Protein expression and crystallization

Recombinant OsWRKY45–DBD (residues 104–182, *O. sativa* Japonica subspecies, GenBank sequence accession AK103959.1) was expressed with the pET22b vector in *Escherichia coli* strain BL21 (DE3). Protein expression was induced by addition of 0.1 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) to log phase cultures grown in LB medium supplemented with 5 μM $ZnCl_2$ and the cells were incubated with shaking at 289 K overnight. Cells were lysed by sonication and 6xHis-tagged OsWRKY45–DBD was collected from the clarified extract with nickel chelating beads (GE Healthcare). The protein was then further purified by gel filtration chromatography on a Superdex75 10/300 GL column (GE Healthcare) equilibrated in 20 mM Tris–HCl, 150 mM NaCl, 5 mM DTT, 5% (v/v) glycerol, and 5 μM $ZnCl_2$ pH 7.3. OsWRKY45–DBD-MUT was expressed and purified by the same procedure for OsWRKY45–DBD.

DNA oligonucleotides used for crystallization were synthesized by Sangon Biotech. DNA duplexes were generated by heating 1:1 mixtures of complementary oligonucleotides at 368 K for 10 min, 337 K for 15 min, and slowly cooling to 289 K. OsWRKY45–DBD was concentrated to 13 mg/ml, mixed with DNA duplexes at 1:1 molar ratio and then incubated on ice overnight to obtain the protein–DNA complex for crystal growth. The crystals for data collection were grown using hanging drop vapour diffusion at 278 K by mixing 0.3 μl complex sample and 0.2 μl reservoir solution containing 0.27 M calcium acetate hydrate and 21% (w/v) PEG 3350.

Data collection, structure determination and analysis

Crystals were cryoprotected in reservoir solution with the addition of 20% (v/v) glycerol and flash frozen in liquid nitrogen before data collection. Diffraction data were collected at the BL-17U beam line of the Shanghai Synchrotron Research Facility (SSRF) and were indexed, integrated and scaled with the program XDS (16). The structure was determined by molecular replacement using Phaser MR implemented in the CCP4 software suite (17) and the AtWRKY1 structure (PDB code: 2AYD) as a search model. Iterative refinement was performed with COOT (18) and PHENIX (19). Data processing and refinement statistics are listed in Supplementary Table S1. All structure figures were made with PyMOL (20). Hydrogen bonds (H-bonds), electrostatic attraction (Arg/Lys side chain nitrogen and phosphate oxygen distance < 5.0 Å) and hydrophobic contacts (C–C distance < 5.0 Å) were determined using the CONTACT routine of the CCP4 software suite. Water mediated interaction refers to H-bonds mediated by one or two water molecules.

Electrophoretic mobility shift assays (EMSA)

In a typical assay, binding reaction with a total volume of 11 μl contained 7.7 μg OsWRKY45–DBD protein and either 2 μg (Figures 1A and 5B) or 4 μg (Figure 5A) dsDNA. The mixture was incubated on ice overnight and the components resolved by 12% non-denaturing polyacrylamide gel electrophoresis (PAGE) in 0.5× TBE buffer at 130 V for 70 min at 289 K. The gels were stained with EB for 10 min, and images were acquired by the fluorescence imaging system.

Circular dichroism (CD) spectroscopy

Protein was prepared in PBS buffer and concentrated to 0.2 mg/ml. The CD spectra were recorded from 200 to 260 nm with the Circular dichroism spectropolarimeter (Chirascan plus, Applied Photophysics Limited, UK) with a cell path length of 1 mm. Each scan was obtained by recording data points at every 0.5 nm with a bandwidth of 2 nm and integration time of 1 s. Final spectra were the average of three

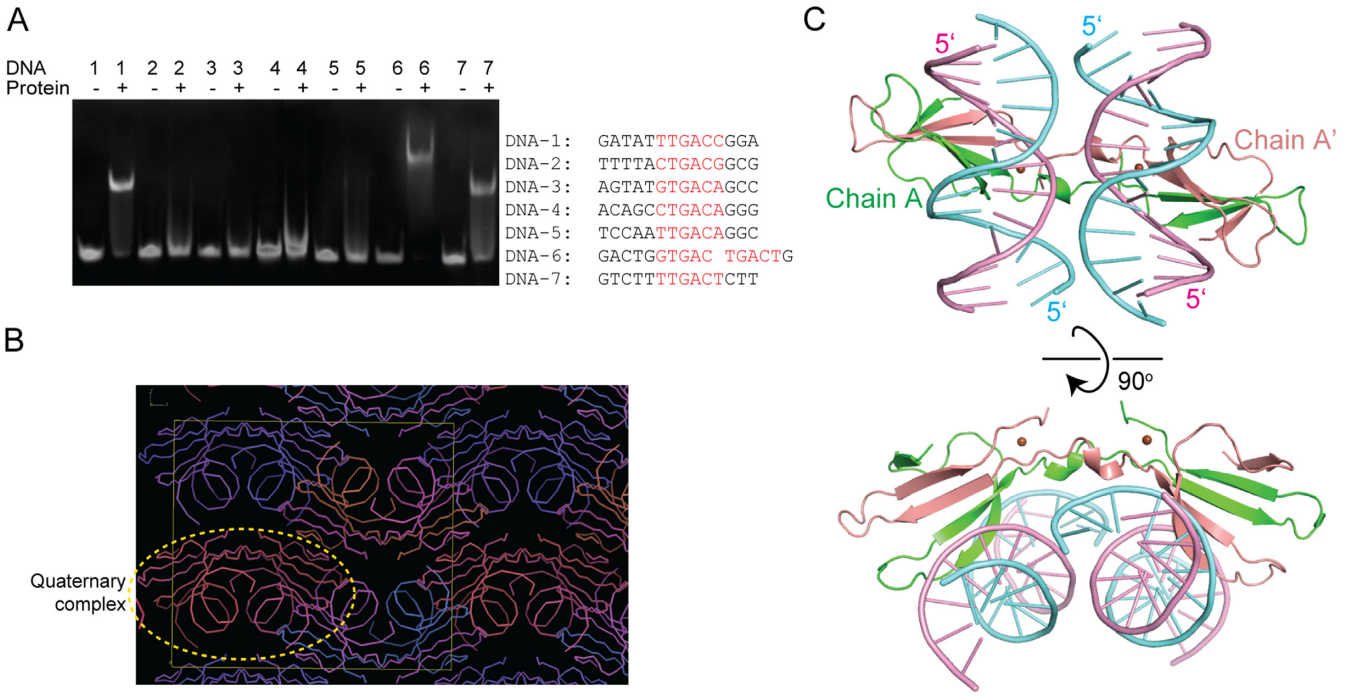

Figure 1. Selection of W-box DNA for co-crystallization and overall structure of the OsWRKY45–DBD/DNA complex. (**A**) EMSA result of OsWRKY45–DBD binding to W-boxes of OsNAC4 promoter. The DNA sequences are shown and W-boxes are labelled red. Note there are two W-boxes in DNA-6. (**B**) Crystal packing of OsWRKY45–DBD/DNA complex. The yellow rectangle denotes the boundary of a crystal cell and the yellow ellipse shows one quaternary complex. (**C**) The OsWRKY45–DBD/DNA complex structure, front view (above) and top view (bottom). Each complex is composed of two OsWRKY45–DBD molecules and two DNA molecules. The two OsWRKY45–DBD molecules are labelled chain A and chain A'. The DNA positive sense strands are coloured aquamarine and the complementary strands are coloured pink. The zinc ions are shown as brown spheres.

scans and corrected by subtracting a spectrum of the buffer recorded under the same conditions.

Clear native PAGE (CN-PAGE)

5 μl of OsWRKY45–DBD (~5.3 mg/ml) was separated by 12% Native-PAGE in TBE buffer. No detergent SDS or anionic dye were used during gel electrophoresis, and therefore this method can be used for separation proteins with pI <7 at physiological pH (21). Electrophoresis was carried out at 289 K at 80 V for 30 min and then 120 V for 90 min. The gel was stained with Coomassie Brilliant Blue R250.

Sedimentation velocity analytical ultracentrifugation (SV-AUC)

Prior to centrifugation OsWRKY45–DBD was concentrated to 0.614 mg/ml (OD_{280} of 0.8) in PBS buffer. Sedimentation velocity experiments were performed in an analytical ultra-centrifuge (ProteomeLab XL-I, Beckman Coulter, USA) at 293 K, in conventional double sector cells with a matched buffer blank in the reference channel at 50 000 rpm in an An50-Ti rotor. Absorbance data (λ = 280 nm) acquired at time intervals of 300 s throughout the run. Data recorded from moving boundaries was analysed in terms of the continuous sedimentation coefficient distribution function $c(s)$ using the program SEDFIT (22–24) where the molar mass (M) of the species in the distribution function is calculated from the fitted Sedimentation coefficient (s) and Stokes radius (R_s) using the equation ($M = sN_A 6\pi\eta^0 R_s/(1-\rho^0 v_{bar})$ which is derived from the Lamm, Svedberg and Stokes-Einstein equations (22–25). Two reference proteins IL-23 (run speed 50000 rpm) and mAb (run speed: 40 000 rpm) were also analysed in the same way to ensure the veracity of the method.

Chemical cross-linking

Glutaraldehyde crosslinks primary amines and was employed to examine the oligomeric state of OsWRKY45–DBD because there are primary amine groups (Lys) on the surface of OsWRKY45–DBD molecules. In a typical crosslinking reaction, a total volume of 20 μl contained 2 μg protein and 0.04% or 0.08% (v/v) glutaraldehyde ($\rho = 0.943$ g/ml), rendering the molar ratio of protein:glutaraldehyde are 4:1 or 2:1, respectively. The reaction was carried out at 293 K for 15 min and quenched by 0.1 M Tris (pH 8.0) for 5 min. The products were fractionated on reducing SDS-PAGE with SurePage™ gels (GenScript) and Tris-MOPS-SDS running buffer (GenScript) and visualised by staining with Coomassie Brilliant Blue R250.

Bio-layer interferometry (BLI) assay

The affinity of OsWRKY45–DBD protein to DNA was determined using Octet RED96 instrument (ForteBio). The experiment was carried out with PBS buffer plus 0.002%

Tween-20 (v/v) at 298 K. In a typical experiment, biotinylated DNA (Sangon Biotech) was immobilized on the surface of streptavidin (SA) sensor for 300 s. The protein was diluted to a serial of concentrations and were loaded onto DNA coated sensor for association for 60 s. Baseline and dissociation were carried out in buffer only for 60 s. The experiment result was analysed by Data Analysis 8.5 software (ForteBio).

RESULTS

Overall structure of the complex

We attempted to express the full-length (residues 1–326) and different truncations of OsWRKY45-1 in *E. coli* and finally chose the 104–182 truncation for subsequent experiments because of its degree of homogeneity observed in both gel-filtration chromatography and SDS-PAGE gel. The 104–182 truncation contains the entire WRKY/DBD domain, so it was named OsWRKY45–DBD. OsNAC4, a positive regulator of programmed cell death associated with the hypersensitive reaction, is downstream of OsWRKY45 and its expression can be upregulated by OsWRKY45. There are seven W-boxes (or W-box cluster) in the OsNAC4 gene promoter, so we synthesized them and tested their binding to OsWRKY45–DBD by EMSA (Figure 1A), and the DNA-1 (5′-GATATTTGACCGGA-3′, the W-box underlined) was chosen for co-crystallization according to its best binding performance to the protein. The 1:1 molar ratio mixture of OsWRKY45–DBD/DNA crystallized at 278 K and the crystals appeared after two weeks.

The crystal of OsWRKY45–DBD/DNA complex diffracted to 2.3 Å (Supplementary Table S1). All the nucleotides of the DNA and residues 110–178 of OsWRKY45–DBD were built in the model. Residues 104–109, 179–182 and the his-tag of OsWRKY45–DBD are invisible in the electron density map due to flexibility or disorder. One asymmetric unit contains one OsWRKY45–DBD molecule, one DNA molecule, and a zinc ion; two OsWRKY45–DBD molecules and two DNA molecules form a quaternary complex as viewed by inspection of the crystal packing (Figure 1B). The oligomeric state of OsWRKY45–DBD is different from the three published WRKY domain structures AtWRKY-1, AtWRKY-4 and AtWRKY52 (RSS1), all of which are monomers. In the OsWRKY45–DBD/DNA complex (Figure 1C), the two DNA molecules are oriented in opposite directions and the angle between them is about 30°, and the OsWRKY45–DBD dimer can be likened to an open clamp that inserts into major grooves of the two DNA molecules.

The OsWRKY45–DBD homodimer

One OsWRKY45–DBD molecule consists of five β strands, numbered β1 to β5 (β1: S111-V114; β2: W123-K129, i.e. the WRKYGQK sequence; β3: K138-C143; β4: Q154-R159; β5: Y167-I172) (Figures 2A, B and 4D). In one OsWRKY45–DBD molecule, the β2–β3 strands and the β4–β5 strands form two hairpins respectively, while the β1 strand stands alone and do not interact with other residues of the molecule; two OsWRKY45–DBD molecules form a dimer by exchanging the β4–β5 strands (Figure 2A). Domain exchanging or swapping, which has been observed in many structures, is the phenomenon that two or more protein monomers exchange structural motifs to form a dimer or oligomer (26), and occurs at a hinge region which is a loop or turn in the monomer but extend after swapping (27). Cartoon model show that β1 to β3 of one chain and β4 to β5 of the other chain forms a five-stranded antiparallel β sheet (Figure 2A and B), which is similar to the AtWRKY4 (PDB code: 1WJ2 and 2LEX, apo and in complex with DNA), AtWRAK1 (PDB code: 2AYD) and AtWRKY52 (PDB code: 5W3X) structures with r.m.s.d. (alignment of the β2–β5 strand) of 1.063, 0.988, 0.688 and 0.709 Å separately (Figure 2C), so we can treat the centrosymmetric OsWRKY45–DBD dimer as two identical antiparallel β sheet domains and a hinge region (between the β3 and β4 strands) in-between (Figure 2A).

One antiparallel β sheet domain of OsWRKY45–DBD dimer contains five β strands, β1–β3 of one chain and β4-β5 of the other chain (named β4′–β5′ for clear reference) (Figure 2B). The long β1–β2 loop strides across the β sheet surface to stabilize the β sheet conformation (Figure 2B). Substantial inter-chain H-bonds are formed between residues of the two chains (Supplementary Table S2), rendering the combined β sheet domain rather stable. The H-bonds are mainly formed between antiparallel β strands, i.e. β1–β5′ and β3–β4′ (Supplementary Table S2). In addition to H-bonds, salt bridges are formed between Lys135 of β2–β3 loop and Asp161 of β4′–β5′ loop (Figure 2B), making the combination of β2–β3 strand and β4′–β5′ strand hard to separate. Notably, the two residues that form the salt bridge are highly conserved in the rice WRKY group IIIa, but absent in other groups (Supplementary Figure S1).

Electron density map of the hinge region between the two β sheet domains is of good quality (Supplementary Figure S2A). Four pairs of H-bonds are formed by residues of the hinge region, which likely stabilize the conformation of the hinge (Figure 2D). Moreover, there is a short α helix (K146-D148) in the hinge region that contributes to the rigidity of the hinge (Figure 2A, B and D). To confirm the existence of the helix, we designed an OsWRKY45–DBD hinge region mutant (OsWRKY45–DBD-MUT) in which the sequence 'THKYDQL' (residues 144–150) was substituted by 'TTPG' which is common in WRKY group I TFs. Analysis by far UV CD spectroscopy reveals the presence of α helix (negative band near 208 nm) in the OsWRKY45–DBD, whilst in the CD spectrum of the OsWRKY45–DBD-MUT this helix signal is much reduced (Figure 2E).

We also investigated the oligomeric state of OsWRKY45–DBD in solution. Initial gel filtration chromatography analysis revealed that OsWRKY45–DBD had an elution volume lower than that expected for the monomer molecular weight (~9.94 kDa) but between that of the 17 and 44 kDa markers, suggesting that OsWRKY45–DBD was multimeric (Supplementary Figure S3A). On CN-PAGE gel, only one main component is observed in the OsWRKY45–DBD protein sample (Figure 2F). However, the band is somewhat smeared and so it is unclear if this represents a monomer, dimer or higher order species in the sample. To address this question, more quantitatively SV-AUC was undertaken. The SV-AUC data

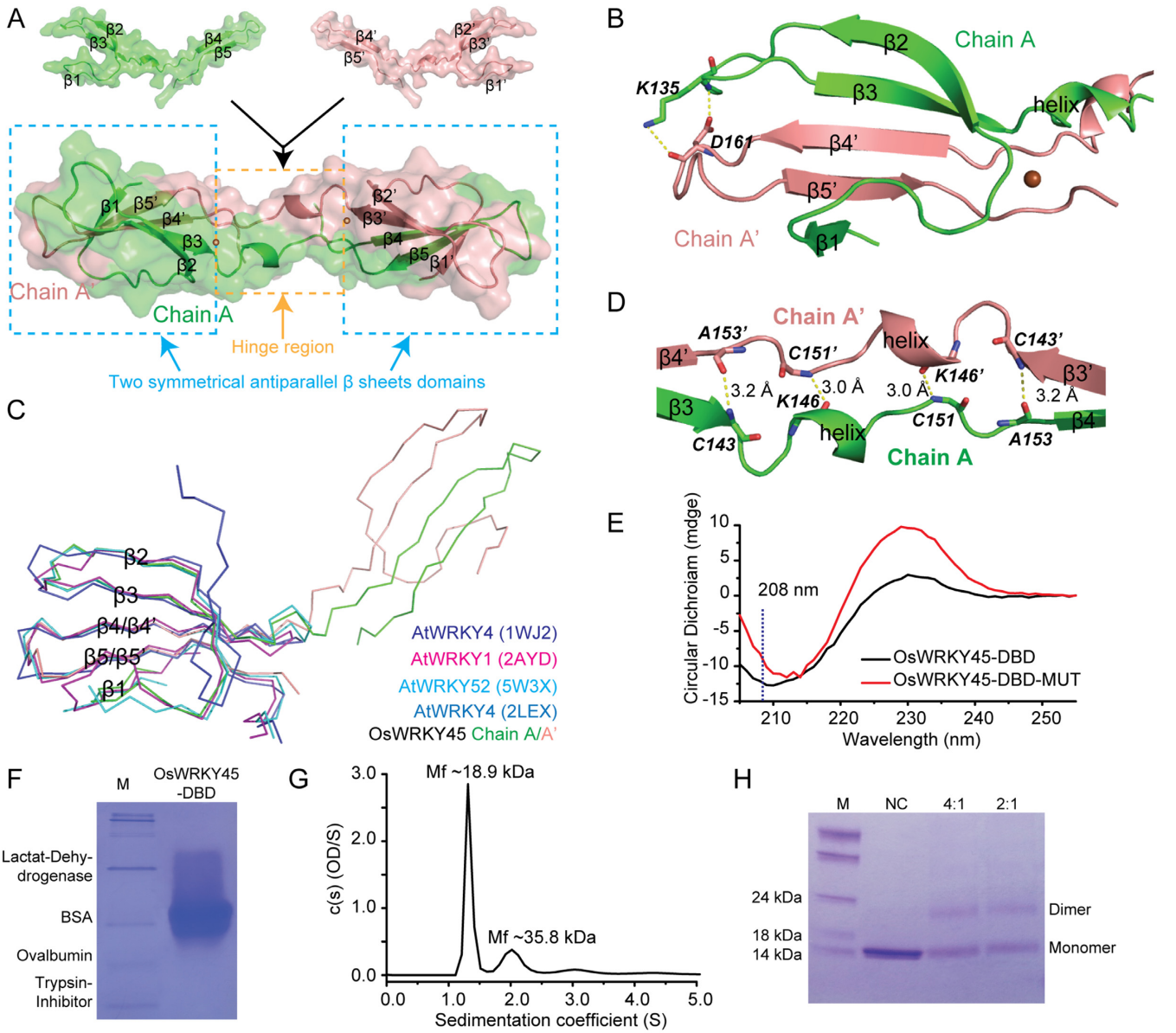

Figure 2. The OsWRKY45–DBD structure. (**A**) Cartoon diagram and transparent surface of the OsWRKY45–DBD monomer and centrosymmetric dimer. The two OsWRKY45–DBD chains are named chain A and chain A' and coloured green and salmon respectively. The β strands of chain A and chain A' are labelled β1–β5 or β1′–β5′. The symmetrical antiparallel β sheet domains and the hinge region in between are boxed blue and orange respectively. The zinc ions are shown as brown spheres. (**B**) Close view of one antiparallel β sheet domain and the hinge region. The two residues K135 and D161 that constitute the salt bridge are shown as sticks (blue for nitrogen and red for oxygen), and the atoms involved are linked by yellow dash lines. (**C**) Structural comparison of one antiparallel β sheet domain of the OsWRKY45–DBD dimer with the other three WRKY domain structures. (**D**) Close view of the hinge region of OsWRKY45–DBD dimer. The inter-chain H-bonds are shown as yellow dash lines, the residues involved are shown as sticks (blue for nitrogen and red for oxygen), and the distance between the hydrogen donors and acceptors are labelled. (**E**) CD spectroscopy assay of OsWRKY45–DBD and its hinge region mutant OsWRKY45–DBD-MUT. The position of 208 nm wavelength are marked as black dash line. (F–H) Analysis of oligomeric state of OsWRKY45–DBD in non-crystallographic conditions. (**F**) CN-PAGE. M: native marker which cannot indicate the molecular weight of a sample. (**G**) SV-AUC. Mf: molecular mass. (**H**) Chemical cross-linking by glutaraldehyde. '4:1' and '2:1': the molar ratio of protein:glutaraldehyde. The dimer is more in the '2:1' lane than the '4:1' lane. M: marker.

was analysed using the *c*(*s*) function that determines the distribution of sedimentation coefficients and frictional ratio that best fit the moving boundary data to derive component molecular weights (see Materials and Methods).These data (Figure 2G, Supplementary Figure S3B) reveal the OsWRKY45–DBD protein sample contains one main component with a sedimentation coefficient of 1.31 S and an associated molecular weight of 18.9 kDa that corresponds to a dimer of OsWRKY45–DBD molecules (2 × 9.94 kDa). There is also a minor component with a sedimentation coefficient of 2.02 S with a molecular weight of 35.8 kDa that could represent a small fraction of OsWRKY45–DBD tetramers in the sample. Analysis of two control samples IL-23 (Interleukin-23, ∼55 kDa) and

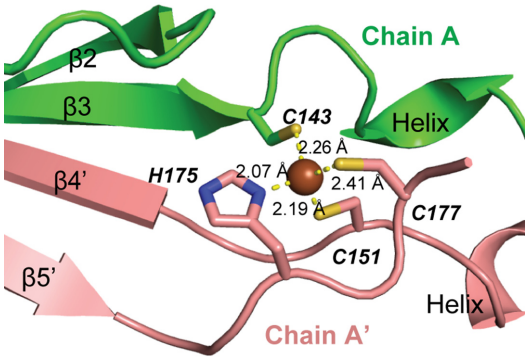

Figure 3. The inter-chain zinc coordination. Coordination of the zinc ion by Cys143 of chain A and Cys151, His175, Cys177 of chain A'. The zinc ion is shown as a brown sphere and the coordinating residues are represented by sticks (blue for nitrogen and yellow for sulphur). The distance between the zinc ion and the coordinate atoms are labelled.

mAb (monoclonal antibody, ∼150 kDa) (Supplementary Figure S3C and D) give the expected molecular weight using this method. Further glutaraldehyde cross-linking experiments were also employed to analyse the solution oligomeric state of OsWRKY45–DBD. These data (Figure 2H) also reveal an intimate association between OsWRKY45–DBD molecules as a dimeric species can be readily captured by crosslinking. Therefore, taken together these experiments strongly support the existence of OsWRKY45–DBD dimers in solution as well as in the crystal structure.

The inter-chain zinc ion coordination

The OsWRKY45–DBD structure belongs to the zinc ribbon category of zinc finger (28). The OsWRKY45–DBD dimer contains two zinc ions contributing two zinc finger motifs. The peculiarity is that the four residues that coordinate with one zinc ion do not come from one chain, but Cys143 from one chain while Cys151, His175 and Cys177 from the other chain (Figure 3). The inter-chain zinc ion coordination or zinc-mediated dimerization/oligomerization mode has also been observed in other proteins such as insulin (29), amyloid-β (Aβ) peptide (30), and glyoxalase-I (31) (Supplementary Figure S4). In the OsWRKY45–DBD structure, the distances between the zinc ion and its coordinating atoms are 2.26 Å (Zn-Cys143-S_G), 2.19 Å (Zn-Cys151-S_G), 2.07 Å (Zn-His175-N_D) and 2.41 Å (Zn-Cys177-S_G), all within the ideal zinc-ligand distances compared with the protein data bank statistical result (32).

Protein/DNA interactions

The OsWRKY45–DBD homodimer uses its positively charged concave side to bind the DNA molecules (Supplementary Figure S5A), and each OsWRKY45–DBD molecule interacts with one DNA molecule with its β2 and β3 strands and β2–β3 loop (Figure 4A). For clear description of the protein–DNA interactions, the DNA sequence (5′-GATATTTGACCGGA-3′, the W-box underlined) of chain B are numbered 1 to 14, making the W-box region 6–11; and the complementary strand (5′-TCCGGTCAAATATC-3′) chain C are numbered 14′ to 1′, making the W-box region 11′-6′ (Figure 4B). The protein–DNA interface can be seen as two part, one is protein with the DNA chain B dA4 to dA9 with a buried area of 292.2 $Å^2$, and the other is protein with the DNA complementary chain C dA6′ to dC12′ with a buried area of 364.6 $Å^2$. The residues of OsWRKY45–DBD and nucleotides of DNA that participate in intermolecular interaction are shown in Figure 4A, and the interaction forces are shown in Figure 4B, C and Supplementary Figure S5B.

Base-specific interactions that confer protein–DNA recognition specificity cover dA4-dA9 and dA6′-dC12′ of DNA, and are mainly focused on dT7, dC8′, dT9′, and dG10′ and partially on dT6 and dG11′ (Figure 4B and Supplementary Figure S5B). Of these base-specific interactions, hydrophobic interaction play important roles as Yamasaki *et al.* reported (14), which mainly include: (a) Lys125 and Tyr126 to dT6 methyl group, (b) Gly127 and Gln128 to dT7 methyl group and carbons of pyrimidine ring, (c) Tyr126, Gly127, Lys129 and Tyr140 to dT9′ methyl group and carbons of pyrimidine ring, (d) Lys129 and Tyr140 to dG10′ carbons of purine ring and (e) Tyr126 and Lys129 to dC8′ and dG11′ carbons of base group rings respectively (Figure 4B). Base-specific interactions also include direct H-bonds (Figure 4B) and indirect H-bonds that mediated by water molecules (33) (Figure 4C and Supplementary Figure S5B). In the crystal structure, cavity is formed in the protein–DNA interface that can accommodate water molecules, and electron density of ordered water molecules are observed that can mediated H-bonds between hydrogen atom donor and acceptor of protein and DNA (Supplementary Figure S2B). Direct H-bonds include: (a) main chain oxygen of Tyr126 to hydroxyl group of dT7, amino group of dA7′ and dC8′, and (b) side chain nitrogen of Lys129 to hydroxyl group of dG10′ (Figure 4B). Water-mediated H-bonds have respect to more residues (Arg124, Lys125, Gln128, Lys129 and Glu130) and nucleotides (dT5, dT6, dG8, dA9, dA6′, dA7′ and dT9′) (Figure 4C and Supplementary Figure S5B).

DNA backbone contacts, i.e. the contacts between protein residues and DNA phosphate and sugar, do not directly contribute to protein–DNA recognition specificity but may play a supporting role. The DNA backbone contacts can be classified into electrostatic attraction, hydrophobic interaction, H-bond, and water-mediated H-bond, and cover dA4-dA7 and dA8′-dC12′ of DNA. Electrostatic interactions (side chain nitrogen of Arg/Lys to phosphate oxygen) include Arg124 to dA4, Lys125 to dT6, and Lys138 to G10′ (Figure 4B). Hydrophobic interaction occur between protein residue carbons and sugar carbons, and include Arg124 to dA4, Lys125 to dT5, Tyr126 to dT9′, Tyr140 to dG10′, Lys129 and Ile131 to dG11′, and Gln132 to dC12′ (Figure 4B). Direct H-bonds are found between protein residues and phosphate oxygen, including main chain nitrogen of Lys125 to dT5, hydroxyl group of Tyr126 to dT9, hydroxyl group of Tyr140 to dG10′, main chain nitrogen of Gln132 to dG11′ and sidechain nitrogen of Gln132 to dC12′ (Figure 4B). Water-mediated H-bonds to DNA phosphate or sugar groups include Trp123, Lys125, Gln128, Glu130, Gln132,

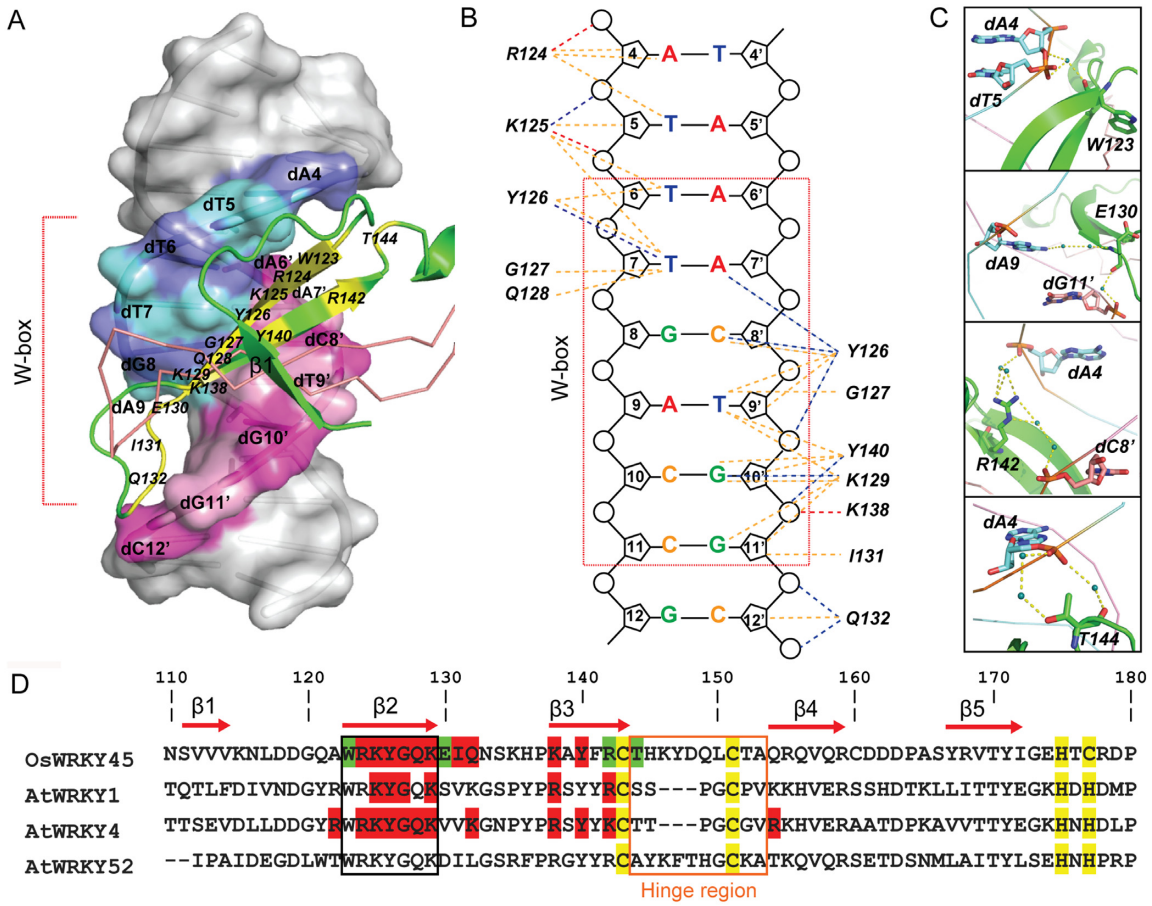

Figure 4. OsWRKY45–DBD/DNA interactions. (**A**) Structural model of the interaction. The OsWRKY45–DBD chain A is shown as cartoon and the chain A' is shown as ribbon, the residues that interact with DNA are highlighted yellow and labelled. DNA is shown as cartoon and transparent surface; the nucleotides that interact with protein are coloured aquamarine/slate for positive-sense strand or magenta/pink for complementary strand and labelled, and those that do not participate in intermolecular interaction are coloured grey. The scope of W-box is shown. (**B**) Detail of interaction forces, including H-bond (blue), electrostatic attraction (red), and hydrophobic interaction (orange). Circle and pentagon represent phosphate group and sugar group respectively; The W-box is boxed red. (**C**) Detail diagram of water mediated H-bond interactions of Trp123, Glu130, Arg142 and Thr144. The residues and nucleotides involved are shown as sticks (blue for nitrogen, red for oxygen, and orange for phosphorus). The water molecules are shown as deep teal spheres. (**D**) Alignment of the three solved WRKY TF structures with CLUSTAL O (Ed. 1.2.4). The conserved cysteines of the consensus zinc finger sequence are highlighted yellow. The residues that have been identified important for DNA interaction are highlighted red or green (participate in water-mediated interactions only). The position of β strands and residue number of OsWRKY45 are shown above. The conserved 'WRKYGQK' sequence and hinge region are boxed black and orange respectively.

Arg142 and Thr144 to dA4-dT7, dC8′, dG11′ and dC12′ (Figure 4C and Supplementary Figure S5B).

Of the OsWRKY45 residues that interact with W-box DNA, Lys125-Glu130 and Tyr140 participate in base specific recognition, and Trp123, Arg124, Ile131, Gln132, Arg142 and Thr144 merely contribute to DNA backbone interactions. All the residues except Glu130 and Gln132 are well conserved through all the groups of WRKY TFs and in accordance with previously reports of the two solved WRKY structures (Figure 4D and Supplementary Figure S1). The water mediated interaction between WRKY and W-box DNA has not been reported previously, and Trp123, Glu130, Arg142 and Thr144 do not have other kinds of interaction with DNA (Figure 4B, C and Supplementary Figure S5B). To confirm the importance of water mediated H-bonds, we measured *in vitro* binding affinity of OsWRKY45–DBD wild-type or R142A mutant to W-box DNA by BLI assay (Supplementary Figure S6A, B). The equilibrium dissociation constant (K_D) measured for OsWRKY45–DBD wild-type binding to W-box DNA is 4.6 μM, while the K_D was 13.8 μM for the R142A mutant, revealing a three-fold reduction in the binding affinity.

Recognition analysis of W-box DNA by OsWRKY45

We have mentioned in the previously part that base-specific interactions mainly focus on dT7, dC8′, dT9′ and dG10′ and partially on dT6 of W-box DNA, thus the middle four nucleotides of W-box contribute to sequence specific recognition by WRKY, and the first and last nucleotides of W-box may play minor roles. To verify our analysis, we designed W-box DNA mutants and evaluated their interaction with OsWRKY45–DBD by EMSA. Mutations of the middle four nucleotides of W-box (MUT7C, MUT8A, MUT9C, and MUT10A) showed no binding to protein (Figure 5A), while MUT6A/G/C and MUT11G/T (partially) retained

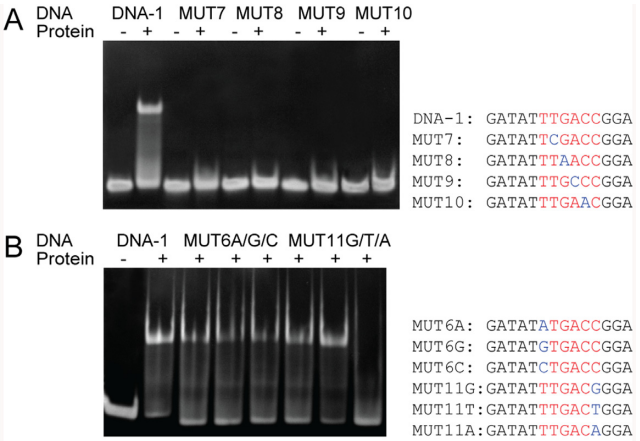

Figure 5. EMSA result of OsWRKY45–DBD binding to W-box DNA mutants of the DNA in the OsWRKY45–DBD/DNA complex structure. (**A**) Mutations of the middle four nucleotides. (**B**) Mutation of the side two nucleotides. The DNA sequence is shown below, the W-box are marked red and the mutated nucleotides are marked blue.

their binding ability (Figure 5B), in consistence with our inference. Since thymine has similar structure as cytosine, the protein binding ability of MUT11T mutant is comparable to wild-type DNA. The MUT11A mutant cannot be recognized by OsWRKY45–DBD, probably because the mutation generated CpA step is not stable and causes change of DNA structure. The pyrimidine-purine (YpR) steps are least stabilized through base stacking interactions, and it has been reported in the DNA binding site of the catabolite activator protein (CAP), two CpA steps caused an overall bending or kink of the DNA of ∼90° (34).

DISCUSSION

Possible mechanism of WRKY domain swapping

Of the four WRKY domain structures solved, AtWRKY1, AtWRKY4 and AtWRKY52 are monomers, while OsWRKY45 is domain-swapped homodimer. It appears that the main driver of WRKY domain swapping is topology of protein which is determined by the primary sequence (35). It has been reported that two loop-deletion mutants of the non-domain-swapping protein monellin have different hinge regions 'YEIKG' and 'YENKG', respectively, and the former can promote domain swapping at the hinge region while the latter cannot, supporting the notion that it is the hinge sequence that determines domain swapping (36). Similarly, a comparison of the four WRKY domain structures shows that the difference arises from the conformation of the hinge region (Figure 2D). In AtWRKY4 and AtWRKY1, the hinge region sequences are 'TTPGCGV' and 'SSPGCPV' respectively (Figure 4D). Here, the proline results in a turn in the main chain (37), and the glycine confers maximal flexibility to allow the hinge region to turn back (38). In AtWRKY52 and OsWRKY45, the hinge region sequences are 'AYKFTHGCKA' and 'THKYDQLCTA' respectively, but the former adopts the monomer conformation as AtWRKY1 and AtWRKY4 while the latter dimerizes (Figure 2D). Structural alignment of their hinges shows that the main chain trend difference starts from the fifth residue (Supplementary Figure S7A). In AtWRKY52, a stable H-bond is formed between the side chains of the fifth residue threonine and the sixth residue histidine and pulls the histidine close to the threonine, and the seventh residue glycine contributes enough flexibility for the main chain to fold back as in AtWRKY1 and AtWRKY4 (Supplementary Figure S7B). However, in OsWRKY45, the fifth residue aspartic acid cannot attract the sixth residue glutamine by constituting a stable H-bond so the glutamine lies more apart from the fifth residue, and the seventh residue leucine has bulky side chain which cannot provide flexibility as glycine (Supplementary Figure S7C). So, we speculate the critical factors that determines the hinge region conformation of WRKY domains of Arabidopsis and rice are: (a) the specific residues proline and glycine that can assist the turn formation; and (b) whether the middle residues of the hinge can form stable H-bond or even salt-bridge. Most group Ia and group II WRKY proteins possess the signature proline and/or glycine, so these WRKYs should be monomer like AtWRKY1 and AtWRKY4 (Supplementary Figure S1A and B). The stable H-bond should involve an acidic or basic amino acid, and checking the middle residues of group Ib and group III WRKY members, the 'TK' (group Ib), 'T/DH' (group IIIa), and 'D/EH/R' (group IIIb) meet the above criteria (Supplementary Figure S1B and C), so these WRKY TFs can form dimers through WRKY domain association. In addition, we mentioned in the introduction section that there are two OsWRKY45 allelic genes in rice, OsWRKY45-1 and OsWRKY45-2, and the difference between their WRKY domains is residue 150 (Leu in OsWRKY45-1 and Met OsWRKY45-2). We measured the K_D of OsWRKY45–DBD L150M mutant (i.e. OsWRKY45-2-DBD) to W-box DNA, which is 7.7 μM and is comparable to OsWRKY45–DBD wild-type (Supplementary Figure S6C).

The zinc ion that bridges dimerization

In the OsWRKY45–DBD homodimer, two zinc ions bridge the two chains by coordinating residues of different chains. One question is why a zinc ion, but not other metal ions, is employed in this role. By aligning the structures of the four WRKY domain structures, we can see the position of the zinc ions and the coordinated residues are coincident (Supplementary Figure S7D). Thus, in essence the zinc-bridged OsWRKY45–DBD homodimer is a zinc finger motif. Maeo *et al.* found only Zn^{2+}, but not other metal ions (Mg^{2+}, Cu^{2+}, Fe^{2+} and Cd^{2+}), can relieve the inhibitory effect of 1,10-*o*-phenanthroline on DNA binding ability of tobacco WRKY protein (39). In the cytosol, the free metal ion concentrations of $Na^+/K^+/Mg^{2+}/Mn^{2+}/Fe^{2+}/Ca^{2+}/Co^{2+}/Ni^{2+}/Zn^{2+}/Cu^+$ (Cu^{2+}) are about $10^{-2}/10^{-1}/10^{-3}/10^{-6}/10^{-6}/10^{-6}/10^{-9}/10^{-9}/10^{-12}/10^{-18}$ (M) (40), but zinc finger proteins have evolved to select zinc ion to stabilize their structures. Factors that determine the selection between metal and protein involve the properties of both metal (valence state, ionic radius, and charge-accepting ability) and protein (charge of residues, dipole moment and polarizability, charge-donating/accepting ability, and number of metal-bound

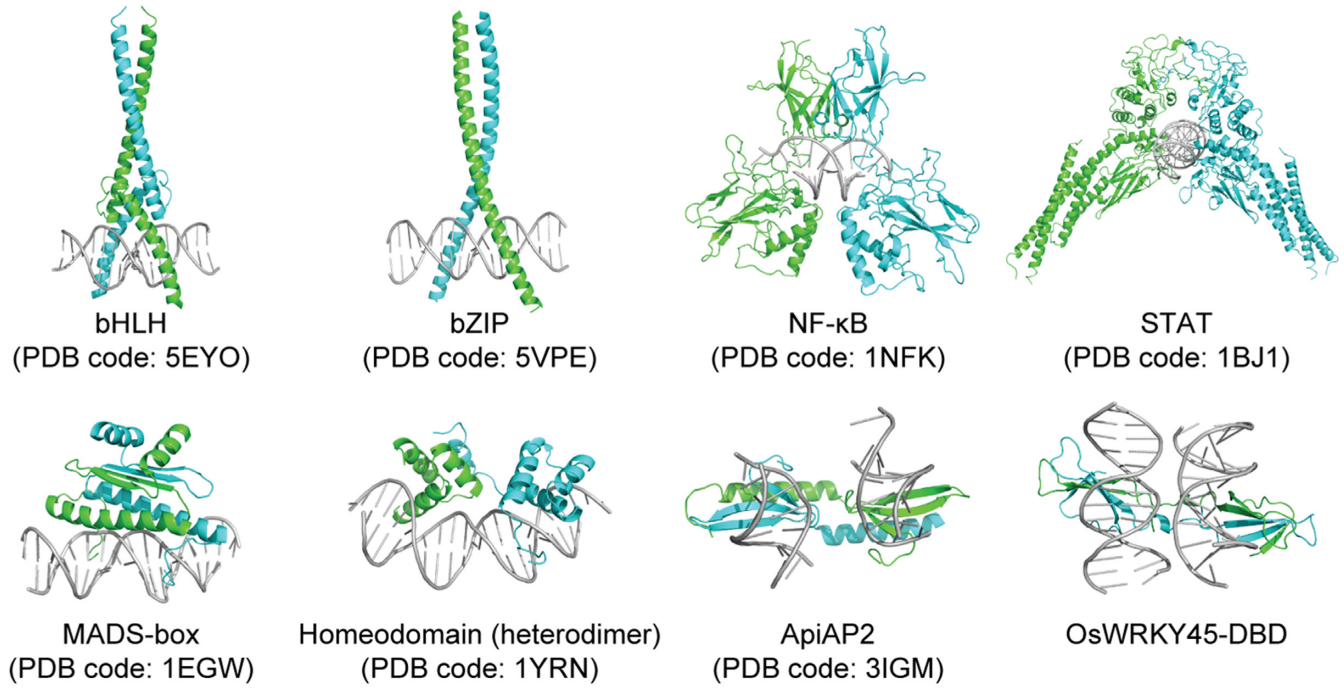

Figure 6. Dimerization modes of TFs on DNA binding. The two monomer of a TF dimer are shown as green or cyan. DNA is shown as grey. bHLH: basic helix-loop-helix; bZIP: basic zipper.

atoms) (40). Irving–Williams series rank the relative stability of complexes formed by divalent metal ions in the following order: $Mg^{2+}<Mn^{2+}<Fe^{2+}<Co^{2+}<Ni^{2+}<Cu^{2+}<Zn^{2+}$, showing that the Zn^{2+} coordinated protein is relatively stable, which is necessary for purposive gene regulation (41).

Biological implication of WRKY domain dimerization

Many eukaryotic TFs contain oligomerization motifs such as basic helix-loop-helix (bHLH) and basic zipper (bZIP) motifs and rely on dimerization to bind DNA (Figure 6) (42). WRKY TFs have also been found to function as dimer or oligomer, e.g. in Arabidopsis, AtWRKY18, AtWRKY40 and AtWRKY60 cooperate by constituting homo- or hetero-complexes, rendering discrepant DNA binding activity and regulating ability (43). It has been proposed that the N-terminal canonical or potential leucine zipper sequences of WRKYs can mediated protein-protein interactions between WRKY members or other proteins (44), and it is peculiar that OsWRKY45 swap the β4-β5 strands of the WRKY domain to constitute a homodimer which is stabilized by zinc ions. The domain swapping of TF dimer to achieve an activated state has also been found in other TFs, e.g. in the p300/CBP (CREB binding protein) reaction intermediate dimer, autoinhibitory loop of one monomer lies near the histone acetyltransferase substrate-binding groove of the other monomer (45). The two DNA molecules in the OsWRKY45–DBD/DNA complex is in opposite direction, which may form a DNA loop to tether two distal DNA loci together to facilitate transcription regulation as ApiAP2 does (Figure 6, Supplementary Figure S8) (46). The DNA-looping mediated by TF binding is common in transcription regulation, such as the 11-zinc finger TF CTCF (CCCTC-binding factor) that assists DNA loop formation by binding to specific recognition sites called CTSs (47), and GATA TFs regulate transcription by recognizing distinct GATA sites with a tandem of two conserved zinc fingers that mediate long-range DNA looping (48). In the OsNAC4 promoter, OsWRKY45 can bind to the DNA-1, six and seven W-boxes, and the DNA-1 W-box is more than one thousand base pairs upstream of the other two, making the formation of a DNA loop by OsWRKY45 dimer binding DNA-1 and DNA-6/7 possible.

The recognition mechanism of W-box by WRKY

The minimal consensus sequence of W-box is TTGACC/T (49,50). Ciolkowski *et al.* investigated the binding of five *Arabidopsis* WRKY proteins to a W-box (AGTTGACC, W-box underlined, other flanking sequences not shown) and its mutants. They found that dG5 (the nucleotide before the W-box) partially determine the binding specificity, and the mutations (m3: ACTTGAC*G*; m4: ACTTGAC*A*; m5: AG*C*TGACC, mutated nucleotide shown as italic) showed reduced or no binding ability to the tested WRKY proteins (50). In the solution structure of AtWRKY4/W-box DNA complex, Yamasaki *et al.* found dC4 and dT5 (the two nucleotides before the W-box) also participate in protein interaction, especially dT5 which is involved in base contact by protein (14). We divide the W-box sequences (including those evaluated in this manuscript and the one in the AtWRKY4/W-box DNA structure) into three groups, fully/partially/no binding, according their binding ability to WRKY TFs (Supplementary Figure S9A). Whether a DNA can be recognized by a protein depends on the se-

quence and shape of the DNA, while the latter can be a consequence of the former to a large extent. Previously we mentioned the impact of YpR steps in the instability of base stacking (34), and the TpA step has the weakest stacking interactions (51), so this can explain why some of the W-boxes lost or reduced WRKY binding ability. The YpR steps are found in the 'no binding' and 'partially binding' groups, including CpA (DNA-3,4,5 and MUT11A), CpG (MUT6A and DNA-2), TpA (MUT11G and DNA-2) and TpG (MUT6G). Of these, the CpA step could cause more serious DNA structure deformation, because DNA-3,4,5 and MUT11A totally lost protein binding ability; while the influence of the other YpR steps is moderate, because MUT6A/G, MUT11G only partially lost binding ability, and DNA-2 need two such YpR steps to totally abolish protein interaction. The reduced binding ability of MUT6C could be due to the replacement of dT6 with dC6 that eliminates the hydrophobic interaction of Lys125 and Tyr126 with dT6 methyl group. In the three W-box mutants (m3, m4 and m5) evaluated by Ciolkowski *et al.*, YpR steps are also present: CpG in m3, CpA in m4, and CpG (5′-GpC-3′) in m5. Based on our complex structure and the analysis above, we can now derive more detailed rules of W-box (numbered 6–11 as above) recognition by WRKY TFs: (a) the middle four nucleotides 7-10 of W-box are irreplaceable; (b) the best starting pair 5 and 6 is TpT, while GpG and TpA/G/C are also acceptable although with reduced affinity to protein; (c) the ending nucleotide 11 should be C/T, guanine is acceptable with reduced protein affinity, while adenine is unacceptable.

Structural view of the OsWRKY45–DBD/DNA binding interface shows that the bases of dT7, dC8′, dT9′, and dG10′ are closer to the surface of OsWRKY45–DBD molecule, while the bases of dT6 and dG11′ are at distal positions from the surface of OsWRKY45–DBD molecule (Supplementary Figure S9B), explaining why the central four nucleotides of W-box determine WRKY domain selection while the two flanking nucleotides plays supporting roles. One additional point is that besides the sequence of W-box, the variance of WRKY domain sequences also contribute to WRKY/W-box recognition specificity. Most of the binding-relevant residue variations are in the first and third residues of the β2–β3 loop (Supplementary Figure S1) that could influence the shape and charge of the DNA binding face of the WRKY domains, and consequently will influence W-box DNA binding preference.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data bank under accession number 6IR8.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all the staff at Shanghai Synchrotron Research Facility (SSRF) beamline BL-17U for help with crystal screening and data collection. We acknowledge the Tsinghua University Branch of China National Center for Protein Sciences Beijing for providing the facility support of AUC. We thank Prof. Xinquan Wang of Tsinghua University for his support on manuscript writing, and Haixia Zhou and Jinfang Yu for helpful discussion. We also thank Jianhui Li of Institute of Biophysics, Chinese Academy of Sciences for his technical help in CD experiment.

FUNDING

National Key Research and Development Program [2016YFD0300700] from the Ministry of Science and Technology, China; 111 Project [B13006] and Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT) Project [IRT1042] from the Ministry of Education, China. Funding for open access charge: National Key Research and Development Program [2016YFD0300700].

Conflict of interest statement. None declared.

REFERENCES

1. Vlot,A.C., Dempsey,D.A. and Klessig,D.F. (2009) Salicylic acid, a multifaceted hormone to combat disease. *Annu. Rev. Phytopathol.*, **47**, 177–206.
2. Cao,H., Bowling,S.A., Gordon,A.S. and Dong,X.N. (1994) Characterization of an Arabidopsis mutant that is nonresponsive to inducers of systemic acquired-resistance. *Plant Cell*, **6**, 1583–1592.
3. Nakayama,A., Fukushima,S., Goto,S., Matsushita,A., Shimono,M., Sugano,S., Jiang,C.J., Akagi,A., Yamazaki,M., Inoue,H. *et al.* (2013) Genome-wide identification of WRKY45-regulated genes that mediate benzothiadiazole-induced defense responses in rice. *BMC Plant Biol.*, **13**, 1471–2229.
4. Sugano,S., Jiang,C.J., Miyazawa,S., Masumoto,C., Yazawa,K., Hayashi,N., Shimono,M., Nakayama,A., Miyao,M. and Takatsuji,H. (2010) Role of OsNPR1 in rice defense program as revealed by genome-wide expression analysis. *Plant Mol. Biol.*, **74**, 549–562.
5. Tao,Z., Liu,H.B., Qiu,D.Y., Zhou,Y., Li,X.H., Xu,C.G. and Wang,S.P. (2009) A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. *Plant Physiol.*, **151**, 936–948.
6. Rushton,P.J., Somssich,I.E., Ringler,P. and Shen,Q.J. (2010) WRKY transcription factors. *Trends Plant Sci.*, **15**, 247–258.
7. Xie,Z., Zhang,Z.L., Zou,X.L., Huang,J., Ruas,P., Thompson,D. and Shen,Q.J. (2005) Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.*, **137**, 176–189.
8. Brand,L.H., Fischer,N.M., Harter,K., Kohlbacher,O. and Wanke,D. (2013) Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and in vitro binding assays. *Nucleic Acids Res.*, **41**, 9764–9778.
9. Ueno,Y., Yoshida,R., Kishi-Kaboshi,M., Matsushita,A., Jiang,C.J., Goto,S., Takahashi,A., Hirochika,H. and Takatsuji,H. (2015) Abiotic stresses antagonize the rice defence pathway through the Tyrosine-Dephosphorylation of OsMPK6. *PLoS Pathog.*, **11**, e1005231.
10. Inoue,H., Hayashi,N., Matsushita,A., Liu,X.Q., Nakayama,A., Sugano,S., Jiang,C.J. and Takatsuji,H. (2013) Blast resistance of CC-NB-LRR protein Pb1 is mediated by WRKY45 through protein-protein interaction. *P. Natl. Acad. Sci. U.S.A.*, **110**, 9577–9582.
11. Fukushima,S., Mori,M., Sugano,S. and Takatsuji,H. (2016) Transcription factor WRKY62 plays a role in pathogen defense and Hypoxia-Responsive gene expression in rice. *Plant Cell Physiol.*, **57**, 2541–2551.
12. Duan,M.R., Nan,J., Liang,Y.H., Mao,P., Lu,L., Li,L., Wei,C., Lai,L., Li,Y. and Su,X.D. (2007) DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein. *Nucleic Acids Res.*, **35**, 1145–1154.

13. Yamasaki,K., Kigawa,T., Inoue,M., Tateno,M., Yamasaki,T., Yabuki,T., Aoki,M., Seki,E., Matsuda,T., Tomo,Y. *et al.* (2005) Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell*, **17**, 944–956.
14. Yamasaki,K., Kigawa,T., Watanabe,S., Inoue,M., Yamasaki,T., Seki,M., Shinozaki,K. and Yokoyama,S. (2012) Structural basis for sequence-specific DNA recognition by an Arabidopsis WRKY transcription factor. *J. Biol. Chem.*, **287**, 7683–7691.
15. Zhang,Z.M., Ma,K.W., Gao,L., Hu,Z., Schwizer,S., Ma,W. and Song,J. (2017) Mechanism of host substrate acetylation by a YopJ family effector. *Nat. Plants*, **3**, 17115.
16. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
17. Potterton,E., Briggs,P., Turkenburg,M. and Dodson,E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 1131–1137.
18. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
19. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
20. DeLano,W.L. (2002) *Pymol Molecular Graphics System*. DeLano Scientific, San Carlos.
21. Wittig,I. and Schagger,H. (2005) Advantages and limitations of clear-native PAGE. *Proteomics*, **5**, 4338–4346.
22. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
23. Brown,P.H. and Schuck,P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.*, **90**, 4651–4661.
24. Schuck,P., Perugini,M.A., Gonzales,N.R., Howlett,G.J. and Schubert,D. (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys. J.*, **82**, 1096–1111.
25. Ebel,C. (2011) Sedimentation velocity to characterize surfactants and solubilized membrane proteins. *Methods*, **54**, 56–66.
26. Bennett,M.J., Choe,S. and Eisenberg,D. (1994) Domain swapping - entangling alliances between proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3127–3131.
27. Mascarenhas,N.M. and Gosavi,S. (2017) Understanding protein domain-swapping using structure-based models of protein folding. *Prog. Biophys. Mol. Biol.*, **128**, 113–120.
28. Gamsjaeger,R., Liew,C.K., Loughlin,F.E., Crossley,M. and Mackay,J.P. (2007) Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.*, **32**, 63–70.
29. Weiss,M.A. and Lawrence,M.C. (2018) A thing of beauty: structure and function of insulin's "aromatic triplet". *Diabetes Obes. Metab.*, **20**(Suppl. 2), 51–63.
30. Polshakov,V.I., Mantsyzov,A.B., Kozin,S.A., Adzhubei,A.A., Zhokhov,S.S., van Beek,W., Kulikova,A.A., Indeykina,M.I., Mitkevich,V.A. and Makarov,A.A. (2017) A binuclear zinc interaction fold discovered in the homodimer of alzheimer's amyloid-beta fragment with Taiwanese mutation D7H. *Angew. Chem.*, **56**, 11734–11739.
31. Cameron,A.D., Olin,B., Ridderstrom,M., Mannervik,B. and Jones,T.A. (1997) Crystal structure of human glyoxalase I–evidence for gene duplication and 3D domain swapping. *EMBO J.*, **16**, 3386–3395.
32. Simonson,T. and Calimet,N. (2002) Cys(x)His(y)-Zn2+ interactions: thiol vs. thiolate coordination. *Proteins*, **49**, 37–48.
33. Joachimiak,A., Haran,T.E. and Sigler,P.B. (1994) Mutagenesis supports water mediated recognition in the trp repressor-operator system. *EMBO J.*, **13**, 367–372.
34. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein–DNA recognition. *Annual Rev. Biochem.*, **79**, 233–269.
35. Ding,F., Prutzman,K.C., Campbell,S.L. and Dokholyan,N.V. (2006) Topological determinants of protein domain swapping. *Structure*, **14**, 5–14.
36. Nandwani,N., Surana,P., Udgaonkar,J.B., Das,R. and Gosavi,S. (2017) Amino-acid composition after loop deletion drives domain swapping. *Protein Sci.*, **26**, 1994–2002.
37. Deber,C.M., Brodsky,B. and Rath,A. (2010) *Proline Residues in Proteins*. John Wiley & Sons, Ltd, Chichester.
38. Marcelino,A.M. and Gierasch,L.M. (2008) Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers*, **89**, 380–391.
39. Maeo,K., Hayashi,S., Kojima-Suzuki,H., Morikami,A. and Nakamura,K. (2001) Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. *Biosci. Biotech. Biochem.*, **65**, 2428–2436.
40. Dudev,T. and Lim,C. (2014) Competition among metal ions for protein binding sites: Determinants of metal ion selectivity in proteins. *Chem. Rev.*, **114**, 538–556.
41. Williams,R.J.P. (2001) Chemical selection of elements by cells. *Coordin. Chem. Rev.*, **216**, 583–595.
42. Amoutzias,G.D., Robertson,D.L., de Peer,Y.V. and Oliver,S.G. (2008) Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.*, **33**, 220–229.
43. Xu,X., Chen,C., Fan,B. and Chen,Z. (2006) Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *Plant cell*, **18**, 1310–1326.
44. Chi,Y., Yang,Y., Zhou,Y., Zhou,J., Fan,B., Yu,J.Q. and Chen,Z. (2013) Protein-protein interactions in the regulation of WRKY transcription factors. *Mol. Plant*, **6**, 287–300.
45. Ortega,E., Rengachari,S., Ibrahim,Z., Hoghoughi,N., Gaucher,J., Holehouse,A.S., Khochbin,S. and Panne,D. (2018) Transcription factor dimerization activates the p300 acetyltransferase. *Nature*, **562**, 538–544.
46. Lindner,S.E., De Silva,E.K., Keck,J.L. and Llinas,M. (2010) Structural determinants of DNA binding by a P. falciparum ApiAP2 transcriptional regulator. *J. Mol. Biol.*, **395**, 558–567.
47. Nagy,G., Czipa,E., Steiner,L., Nagy,T., Pongor,S., Nagy,L. and Barta,E. (2016) Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. *BMC Genomics*, **17**, 637–645.
48. Chen,Y., Bates,D.L., Dey,R., Chen,P.H., Machado,A.C., Laird-Offringa,I.A., Rohs,R. and Chen,L. (2012) DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep.*, **2**, 1197–1206.
49. Rushton,P.J., Torres,J.T., Parniske,M., Wernert,P., Hahlbrock,K. and Somssich,I.E. (1996) Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO J.*, **15**, 5690–5700.
50. Ciolkowski,I., Wanke,D., Birkenbihl,R.P. and Somssich,I.E. (2008) Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol. Biol.*, **68**, 81–92.
51. Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 11163–11168.