

ARTICLE

Open Access

Crystal structures of REF6 and its complex with DNA reveal diverse recognition mechanisms

Zizi Tian¹, Xiaorong Li¹, Min Li¹, Wei Wu¹, Manfeng Zhang¹, Chenjun Tang¹, Zhihui Li¹, Yunlong Liu¹, Zhenhang Chen¹, Meiting Yang¹, Lulu Ma¹, Cody Caba², Yufeng Tong², Hon-Ming Lam³, Shaodong Dai⁴ and Zhongzhou Chen¹

Abstract

Relative of Early Flowering 6 (REF6) is a DNA-sequence-specific H3K27me3/2 demethylase that contains four zinc finger (ZnF) domains and targets several thousand genes in *Arabidopsis thaliana*. The ZnF domains are essential for binding target genes, but the structural basis remains unclear. Here, we determined crystal structures of the ZnF domains and REF6-DNA complex, revealing a unique REF6-family-specific half-cross-braced ZnF (RCZ) domain and two C2H2-type ZnFs. DNA-binding induces a profound conformational change in the hinge region of REF6. Each REF6 recognizes six bases and DNA methylation reduces the binding affinity. Both the acidic region and basic region are important for the self-association of REF6. The REF6 DNA-binding affinity is determined by the sequence-dependent conformations of DNA and also the cooperativity in different target motifs. The conformational plasticity enables REF6 to function as a global transcriptional regulator that directly binds to many diverse genes, revealing the structural basis for the epigenetic modification recognition.

Introduction

Histones are subjected to comprehensive post-translational modifications, including methylation, acetylation, phosphorylation, and ubiquitination. Histone methylation generally occurs on lysine and arginine residues of the N-terminal tails and this can provide an activating or repressive epigenetic mark. H3K27 methylation plays a key role in gene repression and developmental regulation. Most histone demethylases are JmjC-domain-containing proteins and are highly conserved^{1–4}. However, no homologs of metazoan H3K27-specific demethylases KDM6a or KDM6b were found in *Arabidopsis*⁵. Both Relative of Early Flowering 6 (REF6)⁶ and Early Flowering 6 (ELF6)⁷ have been shown to be specific H3K27me3/2 demethylases in plants. JMJ13 is also an H3K27me3 site-specific demethylase, but has no

significant activity on H3K27me2⁸. REF6, ELF6 and JMJ13 are homologous and belong to the KDM4 subfamily. REF6 and ELF6 share a high sequence similarity comprising an N-terminal Jumonji N (JmjN) domain, a JmjC domain, and tandem zinc finger (ZnF) domains in the C-terminus. However, these two demethylases have divergent roles in the regulation of flowering time^{9,10}. Mutants of *elf6* display an earlier floral transition compared with wild type, whereas mutations in *ref6* cause later flowering under different lighting conditions⁹.

Classical C2H2 ZnF domains comprise the largest class of DNA-binding domains¹¹ and are related with various biological processes, including recombination, development, and chromatin regulation. Many histone demethylases contain specialized ZnF domains, such as PHD, Zn-CW, and RING^{1,5} that are involved in histones binding or ubiquitination¹², as well as other functions. However, C2H2 Zinc fingers, which are well established as DNA binding modules, are rarely found in histone demethylases.

The ZnFs of REF6 are found to bind CTCTGYTY (where Y is C or T) DNA motifs directly and recognize

Correspondence: Zhongzhou Chen (chenzhongzhou@cau.edu.cn)

¹State Key Laboratory of Agrobiotechnology and Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Biological Sciences, China Agricultural University, 100193 Beijing, China

²Department of Chemistry and Biochemistry, University of Windsor, Windsor, ON N9B 3P4, Canada

Full list of author information is available at the end of the article.

© The Author(s) 2020



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

In this study, we present crystal structures of the ZnF domains of REF6 and its complex with dsDNA. We identified a novel REF6-family-specific half-cross-braced ZnF (RCZ) domain, a hinge, an acidic region (residues 1224–1239), a basic region (residues 1355–1360), and two classical C2H2 ZnF domains in the structure. Although each REF6 binds directly to six bases, the binding ability and specificity are markedly enhanced by a combination of DNA base and shape readout, and the cooperativity of the binding motifs. Our results reveal a novel mechanism of how the conformational plasticity of DNA enables REF6 to recognize diverse target genes.

Results

REF6 contains an uncommon half-cross-braced ZnF (RCZ) domain

To dissect the structure–function relationship of functionally important domains of REF6, we attempted to express and purify a series of truncated REF6 fragments in *E. coli*. After extensive trials, only one fragment of REF6 (residues 1223–1360) that contains all four zinc fingers (Fig. 1a, b) and is capable of binding the CTCTGYTY motifs produced high-quality crystals at 18 °C. The number and initial positions of the zinc ions (Supplementary Fig. S1) were then determined using the single-wavelength anomalous diffraction (SAD) method by the SHELX C/D/E program¹⁵. The final holo-REF6 structure in the space group $P4_1$ was refined to 1.57 Å resolution (Supplementary Table S1). The structure consists of residues 1238–1353, lacking the N-terminal acidic region and the C-terminal basic region. This is mainly due to conformational flexibility rather than protein degradation during crystallization (Supplementary Fig. S2).

The holo-REF6^{1223–1360} structure contains four C2H2-type ZnF domains. ZnF3 and ZnF4 adopt a fold consisting of a canonical $\beta\beta\alpha$ domain^{16–22} in which a bound zinc ion is sandwiched between the α -helix and the two anti-parallel β -sheets on the other side (Fig. 2a). Unexpectedly, the first two ZnF domains (ZnF1-2) are half-cross-braced. In particular, the Zn1 ion is tetrahedrally coordinated by two cysteines in strands β 1 and β 2, histidine H1263 in helix α 1 from the ZnF1 domain, and histidine H1280 in helix α 2 from the second ZnF2 domain (Fig. 2b). This particular Zn1 ion tightly connects ZnF1 and ZnF2 domains, leading to an overall compact structure (Fig. 2c). We submitted the coordinates of ZnF1-2 (residues 1239–1290) to the Dali server and found no similar structure in the Protein Data Bank (PDB), suggesting that the ZnF1-2 domain forms a novel zinc finger type. Most known cross-braced ZnFs, such as PHD, B-Box, RING, and HIT domains (Supplementary Fig. S3)²³, use two coordinating residues from the first ZnF domain, and two residues from the second ZnF domain, which we denote as 2 + 2 type. The only other type of known half-cross-

braced ZnFs is a cysteine-rich domain of rat brain PKC- γ (Supplementary Fig. S3f)²⁴, which has one coordinating residue from the first ZnF and three from the second ZnF (1 + 3). To the best of our knowledge, the ZnF1-2 domain of REF6 represents a novel discovery of a half-cross-braced ZnF that has a 3 + 1 type coordination, which is distinct from other cross-braced ZnF domains. Therefore, we name this domain as a REF6-family-specific half-Cross-braced ZnF (RCZ) domain.

The complex structure of REF6 with double-stranded DNA

The ZnF domains of REF6 are essential for its recruitment to specific nucleosome sites via direct recognition of the CTCTGYTY motif^{13,14}. To investigate the molecular mechanism of DNA recognition, we first screened a series of double-stranded oligonucleotides ranging from 12 to 22 bp of the *NAC004* (*ATIG02230*) gene, which contains a CTCTGTTT motif, and performed electrophoretic mobility shift assays (EMSA). The EMSA results indicated that the length of the DNA fragments had little effect on the binding (Fig. 3a).

Attempts to cocrystallize REF6 (residues 1223–1360) with different lengths of dsDNA resulted in a complex crystal with a 15-bp DNA fragment (m5). The structure was solved to a resolution of 2.7 Å using the SAD method by SHELX C/D/E (Supplementary Fig. S1d). In the asymmetric unit, one REF6, one dsDNA, two glycerol molecules, and 79 water molecules are modeled into the density map. Compared with holo-REF6, most of the acidic region is resolved in the complex structure. The electron density is well-resolved to allow unambiguous assignment of all the DNA nucleotides (Fig. 3b). Moreover, the composite simulated-annealing sigma-A-weighted $mF_o - DF_c$ omit map clearly confirms the presence of a bound dsDNA (Fig. 3b). Consistent with this observation, the complex was eluted prior to the holo-REF6 which appeared as a single peak following gel filtration chromatography (Supplementary Fig. S4).

dsDNA binding induces profound conformational changes

The binding of REF6 to DNA is important for its *in vivo* demethylase activity^{13,14}. Our complex structure provides details of this interaction at the atomic level. Direct binding with the dsDNA strand involves insertion of the ZnF3 and ZnF4 helices into the major groove. Additionally, the β strands and the zinc ions lie on the outside, forming a tetrahedral structural unit that confers rigidity to the fingers (Fig. 3c). Surprisingly, only helices α 3 and α 4 insert into the DNA major groove (Fig. 3c). Helix α 2 of the RCZ domain is situated above the major groove while helix α 1 projects away from the DNA entirely. Furthermore, the four β -strands of the RCZ domain, with the exception of the C-termini of β 4, also project away from the DNA.

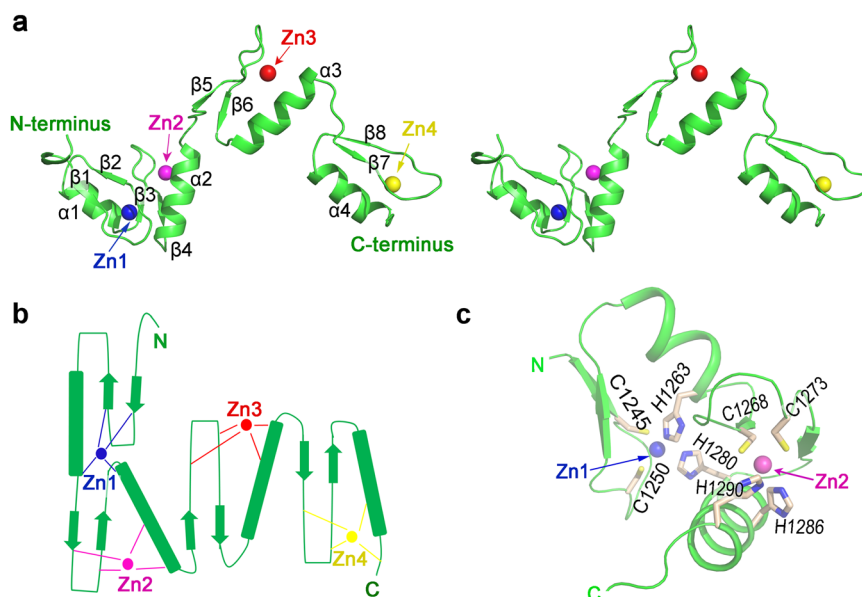


Fig. 2 Crystal structure of the holo-REF6 ZnF domains. **a** Stereoview of the holo-REF6 structure. REF6 is in green color and the four Zn^{2+} ions (Zn1-4) are presented in blue, magenta, red, and yellow spheres, respectively. **b** The topology of REF6 contains a novel RCZ domain and two C2H2-type ZnF domains. **c** Close-up view of the RCZ structure. Residues C1245, C1250, H1263, and H1280 engage the first Zn^{2+} ion (Zn1), while residues C1268, C1273, H1286, and H1290 bind Zn2.

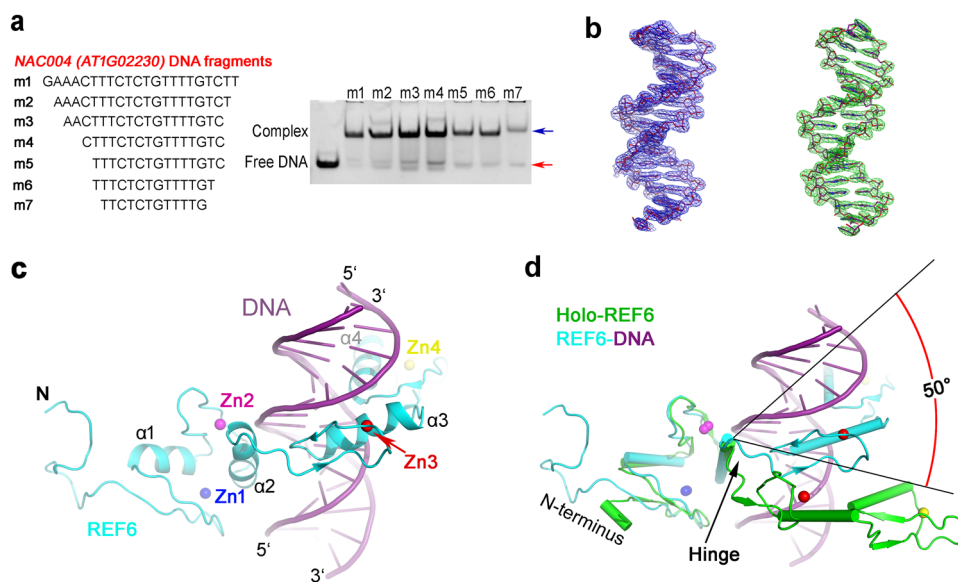


Fig. 3 Structure of the REF6-DNA binary complex. **a** REF6 binds to NAC004 dsDNA of different lengths with the same concentration. Positions of free dsDNA and protein-bound DNA are indicated by red and blue arrows, respectively. **b** The composite simulated-annealing sigma-A-weighted $2mF_o - DF_c$ (left) and sigma-A-weighted $mF_o - DF_c$ (right) electron density maps are shown at 1.5σ and 2.5σ , respectively. **c** Ribbon representation of the REF6-DNA structure. REF6 and DNA are in blue and purple color, respectively. **d** Superposition of the holo-REF6 (green) and the REF6-DNA complex (REF6 in cyan and DNA in purple) structures. The coordinates of the RCZ domains from each were aligned. The last two ZnF domains showed a rotation of $\sim 50^\circ$ towards the dsDNA.

The ZnF1 domain neighbors the acidic region and its surface is less electropositive than the other ZnFs (Supplementary Fig. S5). Moreover, the acidic region repels

DNA. Taken together, this explains why the ZnF1 domain is not directly involved in DNA binding, whereas the ZnF2, ZnF3, and ZnF4 domains are. Most interactions

between the ZnF2 domain and DNA occur on the phosphodiester backbone. Conformations of the holo-REF6 and REF6-DNA binary complex differ greatly with an overall RMSD of 5.3 Å (Fig. 3d and Supplementary Movie S1). Interestingly, the relative position of the ZnF3 and ZnF4 domains (ZnF3-4) does not change during DNA binding and can be viewed as a rigid body, despite the interface between them being only 270 Å². Moreover, the RCZ and ZnF3-4 conformations in these two structures are similar (RMSD = 1.4 and 0.9 Å, respectively). A Dali server search of the ZnF3-4 domain structure identified the ZnF1-2 domain of the Wilms tumor protein (WT1, PDB 6B0R, $Z = 8.3$, RMSD = 2.0 Å)²⁵ and the ZnF1-2 domain of Early growth response protein 1 (Egr1, PDB 1A1F, $Z = 8.1$, RMSD = 1.6 Å)²⁶ as the closest structural homologs (Supplementary Fig. S6a). However, the sequence identity of these domains is less than 30%, indicating low homology. Moreover, only two ZnF domains of REF6 directly insert into the DNA major groove, compared with three or four ZnF domains in Egr1 and WT1, respectively. For the RCZ domain, four sheet-coil transitions (Fig. 1b) and DNA-induced conformational changes of the main chains are observed. Remarkably, structural comparison of holo-REF6 and the REF6-DNA binary complex reveals that the relative orientation of domains RCZ and ZnF3-4 undergoes a rotation of ~50° (Fig. 3d). Therefore, residues 1291–1293 act as a flexible hinge and dsDNA binding induces profound conformational changes of REF6. The shift of the integrated ZnF3-4 domain causes REF6 to form a ring-shaped clamp onto dsDNA (Supplementary Fig. S5a).

Domains ZnF3 and ZnF4 are essential for dsDNA binding

The total buried interface between REF6 and dsDNA reaches 1077 Å², indicating a strong interaction. There are many interactions at the interface between REF6 and dsDNA, including hydrogen bonds, electrostatic interactions, and hydrophobic interactions (Fig. 4a). Most hydrophilic interactions between REF6 and DNA occur on the phosphodiester backbone. The side chains of Lys1275, Tyr1282, His1286, Arg1294, Trp1309, Ser1312, His1316, Tyr1326, Arg1338, Ser1344, Arg1348, and Lys1349, and the backbone of Phe1278, form hydrogen bonds or electrostatic interactions with DNA phosphate groups (Supplementary Fig. S6b). Sequence-independent hydrophobic interactions with the DNA backbone occur via the side chains of Phe1277, Val1289, Phe1307, and Val1340 (Fig. 4a; Supplementary Fig. S6b and Supplementary Table S2). In addition, the side chains of Trp1309, Phe1339, and Val1340 interact with the bases through hydrophobic interactions. For sequence-dependent interactions, the side chain of Tyr1282 interacts with the base of G9 through a hydrogen bond and the base of T10 through hydrophobic interaction. The side

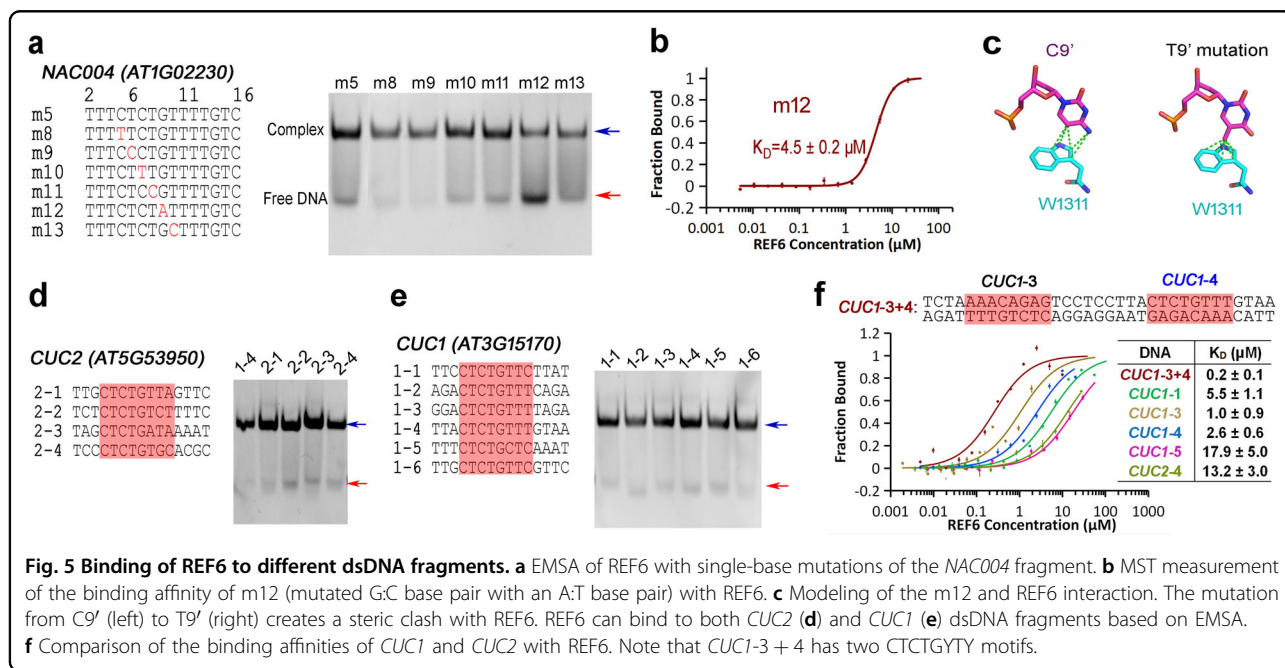
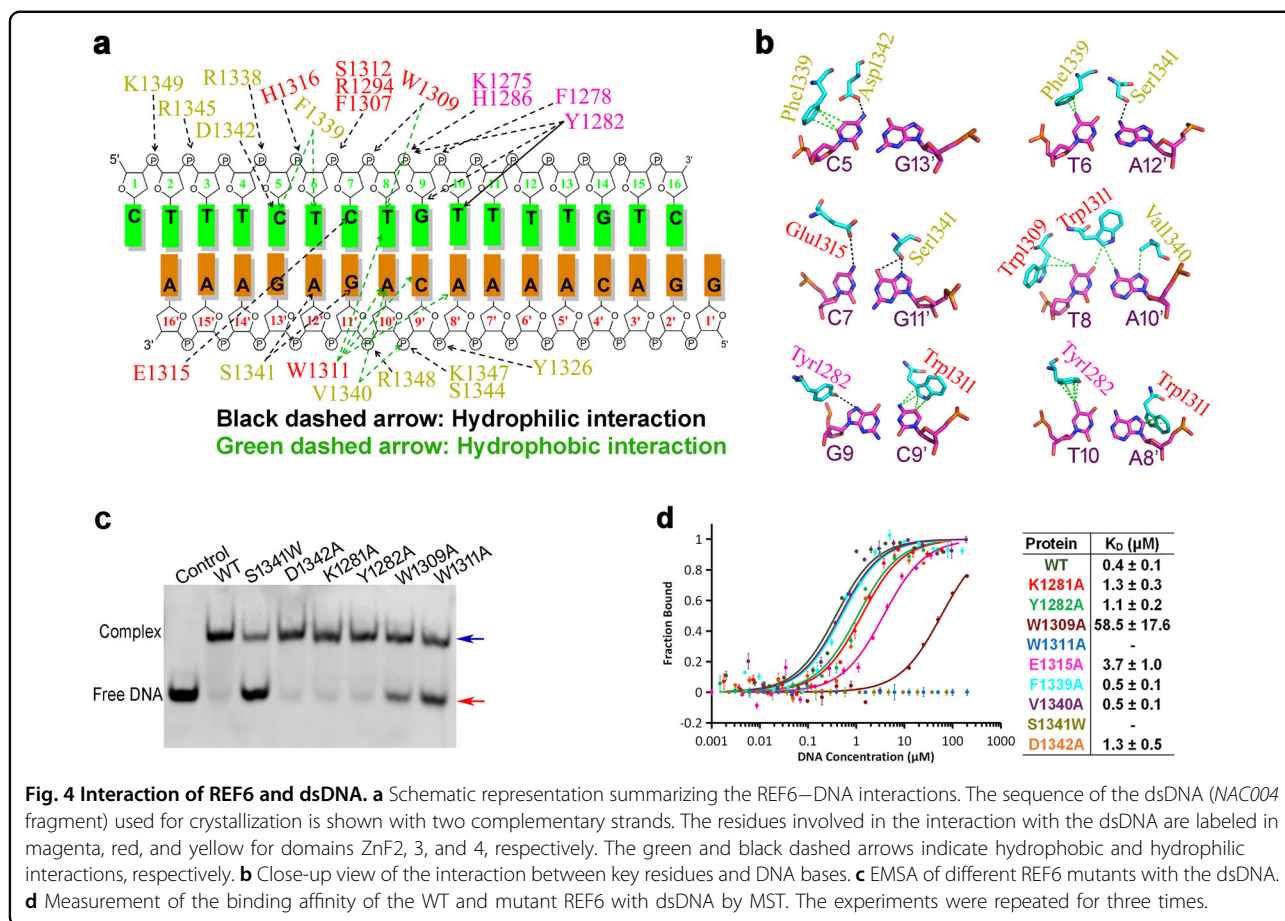
chain of Trp1311 forms a hydrophobic network with the bases of T8, A8', C9', and A10' (where ' indicates the complementary strand). Meanwhile, the residues of Trp1309 and Trp1311 form a hydrophobic environment and together interact with the base of T8. The side chain of Glu1315 forms a hydrogen bond with the bases of C7. The side chain of Ser1341 forms two hydrogen bonds with the bases of G11' and A12'. Moreover, the side chain of Asp1342 interacts with the base of C5 through a hydrogen bond (Fig. 4b).

To test the role of the above key residues, point mutants of REF6 were generated and circular dichroism (CD) experiments were conducted and confirmed that the mutants maintained the same secondary structural elements to that of WT (Supplementary Fig. S7). Meanwhile, EMSAs were performed to understand the qualitative characteristics of DNA binding among the REF6 mutants (Fig. 4c), while microscale thermophoresis (MST) experiments (Fig. 4d) were used to determine the equilibrium dissociation constants (K_D). Consistently, mutants K1281A ($K_D = 1.3 \mu\text{M}$), Y1282A ($K_D = 1.1 \mu\text{M}$), and D1342A ($K_D = 1.3 \mu\text{M}$) exhibited reduced binding affinity to DNA compared with the WT ($K_D = 0.4 \mu\text{M}$), while mutation W1309A ($K_D = 58.5 \mu\text{M}$) exhibited a drastic decrease in binding affinity (Fig. 4d). Mutations W1311A and S1341W were generated in order to reduce or introduce large side chains, respectively. The result of which was the marked abrogation of DNA binding (Fig. 4c, d). Furthermore, we also measured the binding affinity to some weak interactions about the residues E1315A ($K_D = 3.7 \mu\text{M}$), F1339A ($K_D = 0.5 \mu\text{M}$) and V1340A ($K_D = 0.5 \mu\text{M}$). From the MST and EMSA data, mutations of the key residues in ZnF3 and ZnF4 domains dramatically decreased DNA binding. Together, these findings demonstrate that the ZnF3 and ZnF4 domains are essential for dsDNA binding.

DNA methylation reduces the binding of dsDNA to REF6

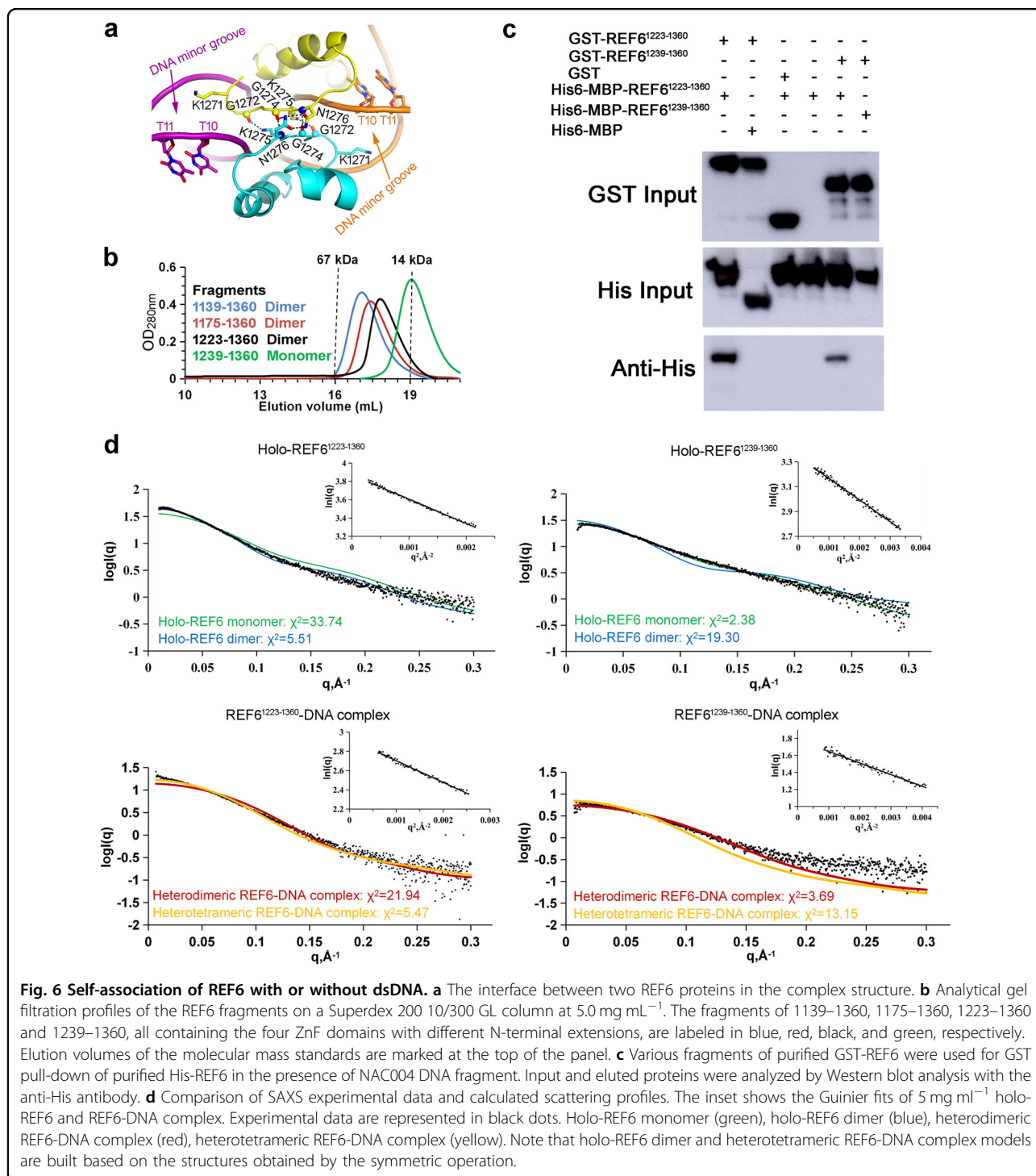
For each strand of dsDNA that contains the CTCTGYTY motif, only five or six bases were directly involved in binding REF6 (Fig. 4a, b). To investigate the sequence specificity of REF6, we first mutated the core DNA sequence CTCTGT, which contributes crucially to REF6 binding (Fig. 5a). The binding affinity of REF6 to the G9A mutant (m12; $K_D = 4.5 \mu\text{M}$) (Fig. 5b) decreased over tenfold compared with the WT fragment (m5; $K_D = 0.4 \mu\text{M}$) (Fig. 4d). Structural analysis revealed that this was mainly due to steric hindrance caused by the C9' T mutation in the complementary strand, which introduced an additional methyl group that clashed with W1311 (Fig. 5c).

The fact that a single-base mutation affects binding suggests that DNA methylation might participate in recognition. Therefore, three different methylated modification sites of the *NAC004* DNA fragments were



synthesized and EMSA experiments were performed. Consistent with the structural observation, the methylated dsDNA of m5-m9C exhibited lower binding affinity than WT (Supplementary Fig. S8a), and the binding affinity of m5-m9C (methylation at 9'C) to REF6 ($K_D = 5.9 \mu\text{M}$) decreased about 15-fold compared with the wild-type m5 fragment ($K_D = 0.4 \mu\text{M}$) (Supplementary Fig. S8b).

Previous studies reported that no REF6 binding signal was detected in *CUC1-1*¹³. To uncover the reason, we analyzed the methylomes of *Arabidopsis thaliana*²⁷. Methylation of this cytosine was found in the CHG methylome. We synthesized the *CUC1-1* dsDNA with and without methylation. EMSA experiments (Supplementary Fig. S8c) confirm that DNA methylation reduces the binding



of dsDNA to REF6 and that the dissociation constant of *CUC1*-1 m⁹C is twice that of the unmodified *CUC1*-1 (Supplementary Fig. S8d). These findings confirm the hypothesis that DNA modification influences the recognition of DNA by REF6.

The shape of DNA affects its binding to REF6

Our structural analyses indicate that only the bases of the CTCTGT motif are directly involved in binding REF6 (Fig. 4a, b). In *Arabidopsis*, *CUC1* and *CUC2*, which are functionally redundant genes, act as key regulators of boundary formation in cotyledons, sepals, and other organs^{28,29}. Previous research revealed that REF6-ZnF domains bound strongly to *CUC1* locus, whereas *CUC2* had no REF6 binding signal¹³. Based on our REF6-DNA complex structure and the above mutagenesis experiments, we speculated that the CTCTGT sequence might be sufficient to bind to REF6. Therefore, we first searched the full-length gene of *CUC2* using the motif CTCTG and identified four sequences, three of which contain the CTCTGT motif (Fig. 5d). Our EMSA experiments showed that REF6 also could directly bind to all four *CUC2* and six *CUC1* motifs in vitro (Fig. 5d, e). To quantitatively compare the difference in binding affinity, we performed MST experiments and found that the affinities for most *CUC1* binding motifs were substantially higher than those for the *CUC2* motifs (Fig. 5f). The MST experiments showed that *CUC1*-5 ($K_D = 17.9 \mu\text{M}$) and *CUC2*-4 ($K_D = 13.2 \mu\text{M}$) exhibited weak binding affinities to REF6. On the other hand, the *CUC1*-3 ($K_D = 1.0 \mu\text{M}$) and *CUC1*-4 ($K_D = 2.6 \mu\text{M}$) motifs had strong REF6 binding affinities (Fig. 5f). Although all of the *CUC1* and binding motifs contained the CTCTGT motif, the affinities differed greatly. We, therefore, hypothesized that sequence-dependent conformations of DNA might confer the observed differences.

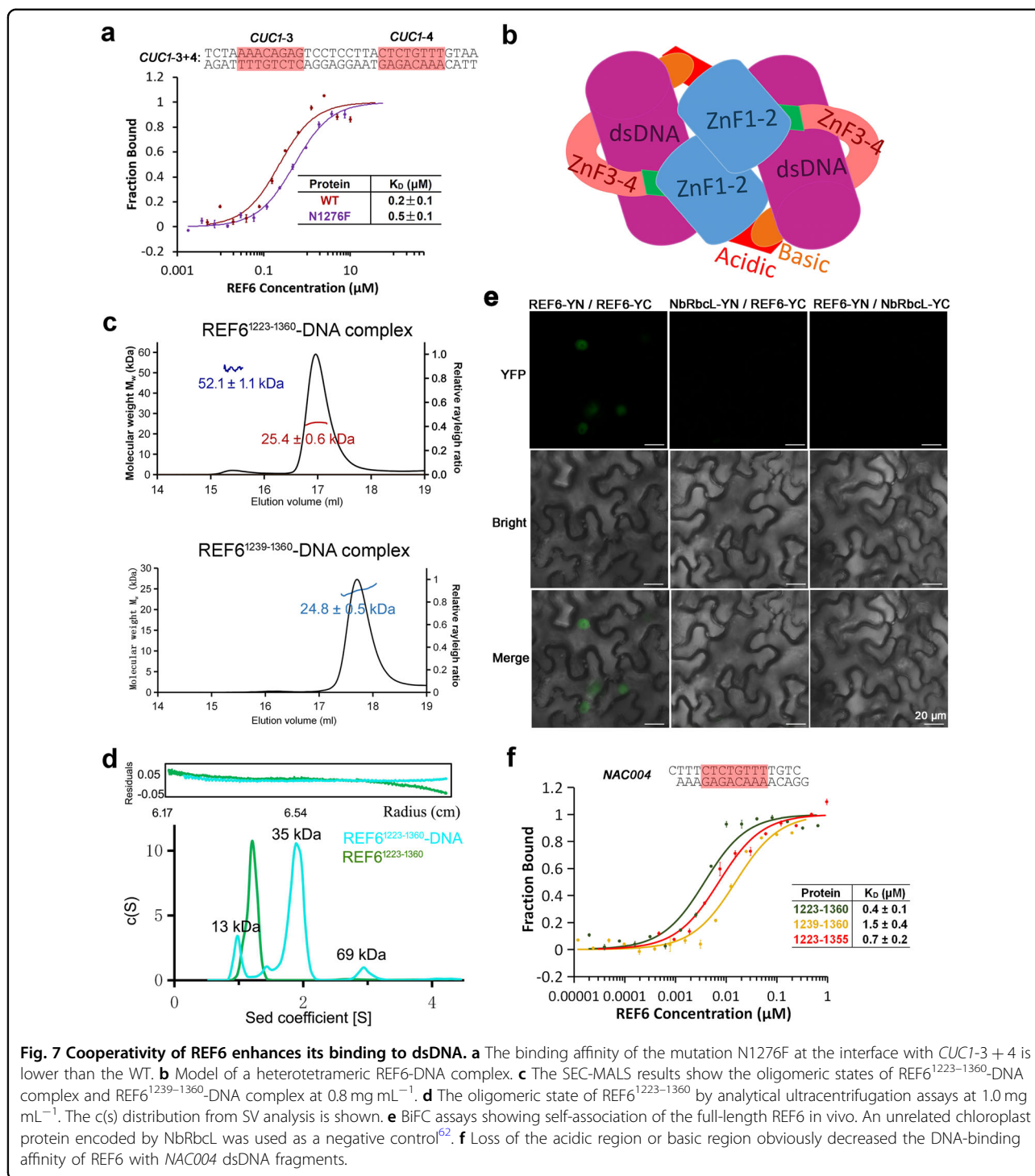
To test this hypothesis, we analyzed the bound DNA conformation in the REF6-DNA complex structure. Structural observation showed that the DNA was bent by 16° compared with the standard B-form DNA duplex (Supplementary Fig. S5b). The minor groove width (MGW) of dsDNA was calculated using the software CURVES+³⁰. The results revealed that the MGW of each nucleotide in the complex structure varied greatly. In particular, the MGW in the middle of T10 and T11 decreased to the lowest value of 2.7 Å, less than half of the standard value of 5.7 Å (Supplementary Fig. S9a, b). The narrower minor groove widened the corresponding major groove and favored the insertion of the helices in REF6 to the major groove. We then calculated the theoretical MGW of dsDNA from *CUC1* and *CUC2* using the DNASHape server³¹. Consistent with the above experimental results, the six CTCTGYTY motifs in *CUC1* except *CUC1*-5 exhibited narrower MGW compared with

the four CTCTG motifs from *CUC2* (Supplementary Fig. S9a). Moreover, further structural analysis revealed that Lys1271 from the neighboring REF6 inserted into the narrow minor groove (Fig. 6a). Consistently, the six motifs in *CUC1*, except *CUC1*-5, exhibited more negative electrostatic potential (EP) compared with the four motifs in *CUC2* (Supplementary Fig. S9b). The more negative EP enhanced the interaction with the positive Lys1271, thereby promoting the binding of REF6. Therefore, DNA shape, including MGW and EP, affects the binding of DNA to REF6.

Self-association of REF6

To study the oligomeric state of holo-REF6 in our crystals, we utilized the Protein Interfaces, Surfaces and Assemblies (PISA) server³². The interaction between molecules from two neighboring symmetric units was not strong, with a buried interface area of 718 Å². To determine the oligomeric state in solution, various REF6 truncations, including fragments 1139–1360, 1175–1360, 1223–1360, and 1239–1360, were generated. According to the elution volume from gel filtration chromatography, fragments of 1139–1360, 1175–1360, and 1223–1360 formed homodimers in solution, but fragment 1239–1360 existed as a monomer (Fig. 6b). Furthermore, GST pull-down experiments showed that GST-REF6^{1223–1360} directly interacted with His6-MBP-REF6^{1223–1360} in vitro, but not with GST or MBP alone (Fig. 6c). Therefore, longer fragments of REF6 interact with each other in solution. To further confirm the oligomeric states of REF6 in solution, small-angle X-ray scattering analysis (SAXS) was performed. As a complementary method to the high-resolution methods of X-ray crystallography, this provides a powerful tool for analyzing dynamic components in solution. We conducted SAXS using a concentration series from 0.5 to 5 mg mL⁻¹, and the results revealed that the oligomerization of REF6 was concentration-dependent. At high concentration such as 5 mg mL⁻¹, the dimeric form fits best for holo-REF6^{1223–1360}, while the monomeric form fits best for holo-REF6^{1239–1360} (Fig. 6d).

In the complex structure, only one REF6-DNA heterodimer (one REF6^{1223–1360} and one m5 dsDNA fragment) was found in the asymmetric unit (Fig. 3c). Because the basic region in the C-termini and some of the residues in the N-terminal acidic region were not visible in the structure, we can not observe the interaction between the C-terminal basic region and the N-terminal acidic region within symmetric units directly. Meanwhile, the interaction between REF6 from two neighboring REF6-DNA complexes occurred between residues 1272–1276 in the ZnF2 domain with a small buried interface area of 234 Å² (Fig. 6a). We analyzed the binding affinity of the mutant N1276F at the interface between two REF6 proteins. The affinity of the double-motif-containing *CUC1*-3 + 4 to



WT was 2.5-fold higher than to the mutant N1276F (Fig. 7a). This result indicated that the residue Asn1276 participated in heterotetramer formation, i.e., the dimerization of two REF6-DNA complexes. The free energy of dissociation (ΔG^{diss}) for the heterotetramer (two REF6-DNA complexes) was $20.3 \text{ kcal mol}^{-1}$, calculated by the PISA server. The large positive value of ΔG^{diss} indicates

that the dimerization of REF6¹²²³⁻¹³⁶⁰-DNA results in a heterotetramer, largely attributed to the dsDNA binding therein. Therefore, we can reasonably construct a model that describes an REF6-DNA heterotetramer by applying a two-fold symmetry operation (Fig. 7b). We further analyzed the oligomeric state of the REF6-DNA complex in solution. The SAXS results implied that the

REF6^{1223–1360}-DNA complex existed remarkably as a heterotetramer, but the heterodimer fits best to the profile for the REF6^{1239–1360}-DNA complex at 5 mg ml⁻¹ (Fig. 6d). Consistent with this observation, the molecular weight of the REF6^{1223–1360}-DNA heterotetrameric complex was obviously larger than the holo-REF6^{1223–1360} dimer (Supplementary Fig. S4). The results from SEC-MALS showed that the REF6^{1223–1360}-DNA complex formed two oligomeric states, the heterodimer and heterotetramer. Meanwhile, the REF6^{1239–1360}-DNA complex only existed as a heterodimer (Fig. 7c). Furthermore, sedimentation velocity (SV) AUC analysis showed that the content of REF6-DNA complex heterotetramer in the REF6^{1223–1360}-DNA complex was significantly higher than those in REF6^{1239–1360}-DNA complex and REF6^{1239–1355}-DNA complex (Fig. 7d and Supplementary Fig. S10). Importantly, chemical crosslinking supports this conclusion (Supplementary Fig. S11).

To further verify the oligomeric state *in vivo*, we performed bimolecular fluorescence complementation (BiFC) assay and confirmed that the full-length REF6 protein interacted with itself. An unrelated chloroplast protein encoded by NbrbcL was used as a negative control and did not interact with REF6 (Fig. 7e). Together, these results suggest that self-association exists for REF6 *in vitro* and *in vivo*.

Cooperativity of REF6 enhances its binding to dsDNA

Neither the N-terminal acidic region nor the C-terminal basic region of REF6 ZnF domains was involved in binding dsDNA directly. However, the loss of either component substantially decreased its DNA-binding affinity. We compared the affinities of different REF6 fragments and found that the affinity of REF6^{1223–1360} was substantially higher than those of REF6^{1239–1360} and REF6^{1223–1355} (Fig. 7f). It is also consistent that the loss of the N-terminal acidic region in REF6^{1239–1360} greatly decreased its interaction with REF6^{1223–1360} (Fig. 6c). Meanwhile, no interaction was observed between GST-REF6^{1239–1360} and His6-MBP-REF6^{1239–1360} *in vitro*. Thus, the N-terminal acidic region is essential for heterotetramer formation. In addition, the mutated residue N1276F at the interface between two REF6 proteins decreased the binding affinity compared with the WT (Fig. 7a). The affinity differences might be due to the missing or mutated residues participating in the formation of the heterotetramer. Thus, the results imply the existence of cooperativity in binding dsDNA and that the self-association of REF6 enhances binding to dsDNA.

Furthermore, we compared the affinities of *CUC1-3*, *CUC1-4*, and *CUC1-3 + 4* which included different numbers of REF6 target motifs. In *CUC1-3 + 4*, the two CTCTGYTY motifs are separated by nine base pairs and are found in the two complementary strands of *CUC1*

(Fig. 5f). The affinity of the double-motif-containing *CUC1-3 + 4* to REF6 was 5- and 13-fold higher than those of *CUC1-3* or *CUC1-4* separately (Fig. 5f), revealing that recognition of the CTCTGYTY motif by REF6 may be cooperatively modulated by the number of neighboring motifs. The cooperative binding of several proteins to one DNA has been reported for *YUC3* by REF6¹⁴ and other proteins³³. Moreover, the large Hill coefficient of 1.9 for the *CUC1-3 + 4* calculated using the Hill model (Supplementary Fig. S12a) further confirmed the cooperativity, while both Hill coefficients were less than 1 for separate experiments using *CUC1-3* or *CUC1-4*. Meanwhile, the cooperativity of REF6^{1223–1360} was higher than that of the mutant N1276F (Supplementary Fig. S12b). These cooperative differences further confirmed the missing or mutated residues participating in the formation of the heterotetramer. Overall, the binding cooperativity may explain the differential binding ability of REF6 to different numbers of CTCTGYTY motifs.

Discussion

The ZnF domains found in histone demethylases are primarily involved in binding histones or ubiquitination. Very few can bind DNA directly (ePHD2 in the PHF6 protein³⁴, the CXXC domain of JHDm1 subfamily³⁵, and the ZnF domains in the REF6 protein subfamily). To the best of our knowledge, reports on complex structures of C2H2 Zinc fingers with DNA are rare for histone demethylases. Here we present a new mechanism of histone demethylase recruitment to specific chromatin sequences. These results provide structural insights into the recognition mechanism of REF6 and its cognate DNA motif, including that the bases and DNA backbones interact with different ZnF domains to further enhance binding specificity. The flexible hinge region allows for reorientation of the ZnF3-4 domain in order to facilitate insertion into DNA (Fig. 3d), resulting in a more compact conformation compared with that of holo-REF6 (Supplementary Fig. S5a).

We show that the RCZ domain adopts a novel fold of the half-cross-braced ZnF protein family. All other known cross-braced ZnFs have two or three coordinating residues cross-braced from the second ZnF domain^{23,24}. In the RCZ domain, only one residue is cross-braced. This finding may be connected with the specific recognition mechanism of the REF6 family. Sequence alignment from different plant species shows that the residues in the C-termini of the ZnF2 domain are conserved (Fig. 1b). The conservation suggests their functional importance and is consistent with our structural findings. For the RCZ domain in the DNA-bound structure, only the C-termini of β_4 and α_2 bind the phosphodiester backbone. Thus, the RCZ domain serves to stabilize DNA binding, but does not function in base recognition. Moreover, compared

with the ZnF3-4 domain, the RCZ domain undergoes a remarkable conformational change upon binding (Fig. 3d) that involves four sheet-coil transitions (Fig. 1b) and a rotation of the N-terminus. Furthermore, the interaction residues between two neighboring REF6 proteins in the solved complex structure, as well as Lys1271 which can insert into the narrow minor groove, are all from the RCZ domain. This reveals an important regulatory function in binding DNA. Phylogenetic analysis shows that the RCZ domain might exist in plants only, suggesting a lineage-specific role.

Based on our structures, only the three-base CTC motif is directly recognized by the ZnF4 domain of REF6. Like ZnF4, the ZnF3 domain is positioned in the DNA major groove. It improves the binding affinity via many hydrophilic interactions with the phosphodiester backbone and hydrophobic interactions between Trp1309 and Trp1311 with nucleic acid bases. A previous study demonstrated that REF6 bound to *CUC1* in vivo, with the exception of *CUC1-1*, but not *CUC2*¹³. However, our in vitro experiments confirmed that four CTCTG motifs from *CUC2* bound REF6 directly. Moreover, the accurate binding experiments further showed that the affinity of all six *CUC1* and four *CUC2* motifs differ greatly (Fig. 5f), revealing the diversity of binding motifs. This raises the question of how REF6 achieves its substrate specificity in vivo due to the presence of 10 million such motifs in the *Arabidopsis* genome. When the oligonucleotide length is shorter than 12 bp, the binding affinity of REF6 decreases dramatically. For example, the binding affinity of the 10-bp sequence TTCTCTGTTT ($K_D = 483 \mu\text{M}$), which contains all of the interactions found in the structure, is very low compared with the 16-bp motif ($K_D = 0.4 \mu\text{M}$), revealing that DNA conformations are involved in DNA recognition.

Several studies have suggested that DNA shape, including MGW and electrostatic potential, contributes to specific DNA recognition^{33,36,37}. Our experiments show that the CTCTGTTTT motif has the highest binding affinity, while other CTCTGYTY motifs are weaker. Thus, our results can explain the finding that only 15% of CTCTGYTY motifs are within REF6 target sites¹³. The theoretical MGW and electrostatic potential reveal that *CUC1-3* and *NAC004* DNA, both with a TTTT sequence 3' to CTCTG, have the narrowest minor groove and the most negative electrostatic potential (Supplementary Fig. S9) in the middle of T10 and T11. Meanwhile, the dissociation constant of REF6 towards the T-tract mutant (m13; $K_D = 10.6 \mu\text{M}$; Supplementary Fig. S9d) decreased over 25-fold compared with the WT m5 fragment ($K_D = 0.4 \mu\text{M}$). These results are consistent with the report that T-tract sequences tend to form a narrow minor groove^{38,39}. Therefore, the narrower minor groove is an intrinsic structural feature of the DNA sequence. In

addition, the feature may become further enhanced by the binding of REF6 as the value in the complex structure is smaller than the theoretical value (Supplementary Fig. S9a). Except for the role in widening the corresponding major groove and favoring the insertion of the REF6 helices, the specific DNA shape will favor the binding of Lys1271 using its more negative electrostatic potential. Thus, we suggest that shape readout of DNA structure may be a hallmark of the REF6 protein subfamily and constitutes a mechanism for REF6's selectivity towards CTCTGYTY motifs.

Interestingly, we found that REF6 displays positive cooperativity. In addition, the ZnF3 domain might mediate self-association similar to the closest structural homolog, the ZnF1 in the Wilms tumor protein²⁵ (Supplementary Fig. S6a). In both the holo- and DNA-bound REF6 structures, only one REF6 molecule was solved. However, the free energy of assembly, in vitro, and in vivo experimental results showed the existence of dimerization between REF6 fragments and heterotetramer formation with bound DNA. The acidic region and basic region are found at both termini of REF6 ZnF domains (Fig. 1a). The basic region is especially conserved in many plant species (Fig. 1b), revealing its functional importance. The acidic region appears to be nonconserved, but many continuous aspartic acid or glutamic acid residues are also found in the upstream region of the ZnF domains in many plant species. Although it is not resolved in the structures, the basic region was speculated to interact with the acidic region of a neighboring REF6 molecule (Fig. 7b). Missing or mutated residues participating in the formation of the heterotetramer substantially decreased its DNA-binding affinity (Figs. 6c, 7a). Moreover, the affinity of the double-motif-containing *CUC1-3 + 4* was dramatically higher than those of the *CUC1-3* or *CUC1-4* fragments separately (Fig. 5f). Additionally, the shape of the binding curve is steeper for the longer construct compared with the short one (Supplementary Fig. S12). These results suggest the existence of cooperativity in binding dsDNA and that the self-association of REF6 greatly enhances the affinity. Thus, the REF6-DNA complex is able to form a heterotetramer, but the ratio is low. As one REF6 molecule interacts with dsDNA by forming a semi-circular structure across ZnF2-4 (Fig. 3c), two REF6-dsDNA complexes may dimerize to form a heterotetramer that encircles the oligonucleotides completely (Fig. 7b). The result of which substantially enhanced the binding stability. This cooperativity can explain the fact that REF6 tends to bind CTCTGYTY-motif clusters¹³.

DNA methylation is one of the most well-studied epigenetic marks. From our structures and in vitro experiments, we find that DNA methylation at specific sites disfavors the binding of REF6 to DNA. Our discovery of *CUC1-1* methylation in the methylomes of *Arabidopsis*

explains a previous report that no REF6 binding is detected when using a *CUC1-1* substrate¹³. Recently, it is also reported that other factors, such as chromosome states, binding partners, the number of binding motifs, and even dependence on transcription factors, may influence the recognition by REF6 in vivo^{40–44}. Take the case of CTCTGYTY-motif clusters, due to the high affinity from cooperativity, REF6 can tolerate sequence variation in the CTCTGYTY motif and thus might bind many kinds of motifs. All of the features found in our study, such as DNA methylation, MGW, and cooperativity, can contribute to the recognition of diverse sequence motifs. Thus, our results explain the finding that about 80% of REF6-targeted DNAs contain the CTCTGYTY motif and only 15% of CTCTGYTY motifs in the *Arabidopsis* genome overlapped with REF6-binding sites¹³.

In summary, these results uncover new recognition and recruitment mechanisms for histone demethylation. Despite the similarity of the REF6 residues at the interface compared with the complex structure of REF6 ZnF2-4 with dsDNAs⁴⁵ (Supplementary Fig. S13), our structures and data reveal that REF6 recognizes dsDNA through diverse mechanisms, such as base and shape readouts, cooperativity of protein and binding motifs, and DNA methylation. Our findings not only provide logical explanations for the selective recognition of selected CTCTGYTY motifs by REF6, but also reveal the crosstalk between DNA modification and histone modification. Furthermore, the high affinity of the REF6 tandem ZnF domains in binding DNA, and/or in combinations with other zinc finger motifs, might be used as a tool for specific gene regulation interventions under different cell contexts or stimuli, such as curing or preventing diseases including cancers, diabetes, cardiovascular diseases, and viral infections.

Notes on the complex structures of REF6 ZnF2-4 with DNAs

When we were preparing our manuscript, another group released complex structures of REF6 ZnF2-4 with methylated dsDNAs in the PDB (6JNN, 6JNM, and 6JNL)⁴⁵. A quick comparison with our structures indicates that the binding interface between REF6 and DNA bases are similar. However, the sequence lengths of protein and nucleic acid in our complex structures are significantly longer. Thus, the narrow minor groove, the shape readout mechanism, and the cooperativity of REF6 were not observed in those structures.

Furthermore, our structures reveal that the ZnF1 is critical in stabilizing the conformation of the ZnF1–2 and important for the self-association and synergy of DNA recognition. Moreover, no holo-REF6 structures were reported. Our holo-REF6 structure also revealed a large

conformational change required for the ZnF3-4 to bind DNAs. Thus, the major conclusions in our manuscript are not obtained in those structures. We believe our structures are more physiologically relevant than the aforementioned ZnF2-4 structures.

Materials and methods

Cloning, protein expression, and purification of REF6

Constructs of *Arabidopsis thaliana* REF6 were subcloned into either a modified pET28a vector encoding a His6-MBP fusion tag or the pGEX-4T-2 expression vector^{33,46}. Both the MBP and the GST fusion tags are removed from the REF6 constructs by a tobacco etch virus (TEV) protease cleavage site. All mutants of REF6 were created using site-directed mutagenesis and verified by DNA sequencing. The plasmids were transformed into *Escherichia coli* BL21 (DE3) cells. Cells were grown in Luria-Bertani (LB) medium at 37 °C until OD₆₀₀ reached 0.8–1.0 and the culture was then supplemented with 100 μM ZnCl₂, and protein overexpression induced at 18 °C by 0.2 mM isopropyl β-D-1-thiogalactopyranoside. After 12 h, cells were harvested by centrifugation at 4000 × *g* for 10 min at 4 °C, and then resuspended in a lysis buffer (20 mM Tris pH 7.3, 1 M NaCl, 2 mM β-mercaptoethanol), replenished with 0.1% (v/v) Triton X-100 and 1 mM Phenylmethanesulfonyl fluoride (PMSF) (Invitrogen). Cells were lysed by sonication and clarified by centrifugation at 18,300 × *g* for 45 min. The lysate was filtered through a 0.45 μm filter membrane to remove cell debris before being loaded onto a Ni²⁺-chelating column or GST affinity column (GE Healthcare). The fusion tag was cleaved using TEV protease at 4 °C for 6–8 h when necessary. The proteins were further purified by size exclusion chromatography. The peak fractions were collected and pooled, and the purity was assessed by SDS-PAGE. The proteins were concentrated to 10–15 mg mL⁻¹ before setting up for crystallization. The REF6-DNA complex was obtained by directly incubating the REF6 protein with an excess (1:1.1) of the specific DNA substrate in 20 mM MES (pH 6.0), 150 mM NaCl, 2 mM β-mercaptoethanol for 2 h on ice. The complex was further concentrated to 6 mg mL⁻¹ for crystallization.

Crystallization and data collection

Crystallization trials of the holo-REF6 and the REF6-DNA complex were performed with multiple commercial crystallization kits in 48-well plates using the hanging drop vapor diffusion method. The DNA oligonucleotides were prepared by annealing two complementary oligonucleotides (5'-CTTTCTCTGTTTTGTC-3' and GGA CAAAACAGAGAAA). Briefly, a mixture of the two oligos was incubated at 95 °C for 5 min and then allowed to cool down slowly to 20 °C in 2 h. The 15-bp duplex DNA with single-nucleotide overhangs at the 5'-ends was added

in three-fold molar excess to 368 μM (6 mg mL⁻¹) of REF6. High-quality holo-REF6 crystals were grown in a reservoir solution containing 12% PEG 20000, 100 mM MES, pH 6.5, at 18 °C, and the complex crystals were obtained in 100 mM HEPES (N-(2-hydroxyethyl) piperazine-N'-(2-ethanesulfonic acid)), pH 7.5, 26% PEG 400, and 100 mM CaCl₂ at 18 °C. All crystals were cryo-protected in the mother liquor supplemented with 25% (v/v) glycerol and flash-frozen in liquid nitrogen before data collection. The data were collected on beamlines BL17U1 and BL19U1 at the Shanghai Synchrotron Radiation Facility (SSRF). Data collection statistics are summarized in Supplementary Table S1.

Structure determination

Structures of holo-REF6 and REF6-DNA complex were both solved by using the SAD method. The anomalous signals in the data were strong as analyzed by the SHELX C program¹⁵, indicating the existence of zinc atoms. For holo-REF6 in space group *P*4₁, four initial Zn sites were found by the program SHELX D with a $CC_{\text{weak}}/CC_{\text{all}}$ of 26.5/43. A crude partial model with four α -helices and eight β -strands in 102 residues was generated by SHELX E¹⁵ and the figure of merit reached 0.706 with a CC_{part} of 48.55%. The model was further extended using AUTO-BUILD⁴⁷ followed by manual model building using COOT⁴⁸ and refinement using REFMAC5⁴⁹ iteratively. The native dataset of the crystal grown at 18 °C was then used and the structure was refined finally to 1.57 Å resolution with an R_{work} of 21.1% and an R_{free} of 23.4%. For the REF6-DNA complex, we failed to find the phase with molecular replacement using the holo-REF6 coordinate. The complex structure was then solved by the SAD method using the anomalous signals of zinc atoms. The REF6-DNA complex crystal was grown in space group *P*3₂21. Four initial Zn sites were found by SHELX D. A partial model with three α -helices and four β -sheets in 89 residues was generated by SHELX E¹⁵ and the figure of merit reached 0.695 with a CC_{part} of 22.62%. The model was further built in program Buccaneer⁵⁰ to an $R_{\text{work}}/R_{\text{free}}$ of 0.41/0.47. The primitive density map showed a clear outline of nucleic acids. At this point, the standard structure of a B-form nucleic acid was fitted into the electron density map. After iterative manual model building in COOT and refinement with REFMAC5, the structure was refined to 2.7 Å resolution with an R_{work} of 22.1% and R_{free} of 25.7%.

All structural figures in this article were prepared using PyMOL (version 1.8.0.0, Schrödinger LLC)⁵¹.

Gel filtration chromatography

Purified recombinant REF6 proteins were applied to gel filtration columns (Superdex-200, GE Healthcare) equilibrated with a buffer containing 20 mM HEPES, pH 7.3,

500 mM NaCl, and 2 mM β -mercaptoethanol. Peak fractions were collected and visualized by SDS-PAGE followed by Coomassie Bright Blue staining.

Electrophoretic mobility shift assays

Double-stranded DNAs used for EMSA were prepared by annealing two complementary oligonucleotides. Briefly, a mixture of the two oligos was incubated at 95 °C for 5 min and then allowed to cool down slowly to 20 °C in 2 h. About 30 μM protein was incubated with 10 μM dsDNA in a 20 μL reaction mixture (containing 20 mM Tris pH 8.0, 100 mM NaCl, 2.5 mM MgCl₂, 0.1% CA-630, 10% glycerol, 1 μM ZnSO₄, 1 mM Dithiothreitol (DTT))¹³ for 1 h at 4 °C and 4 μL of the mixture was then separated on a 10% native polyacrylamide gel in 0.5 \times TBE buffer (45 mM Tris, pH 8.0, 45 mM boric acid, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0) at 180 V for about 30 min. The DNAs were visualized by staining with ethidium bromide.

Microscale thermophoresis

The binding affinities between dsDNA oligonucleotides and REF6 were measured using microscale thermophoresis (MST). The oligonucleotides were labeled with 6-carboxyfluorescein (FAM) and annealed slowly from 95 to 20 °C similar to the oligos used for cocrystallization. Wild-type and mutated REF6 proteins were mixed with dsDNA in a buffer containing 20 mM HEPES, pH 7.3, 150 mM NaCl, 0.2 mM TCEP and 0.5% CA-630 (Sigma: 18896). After incubation for 30 min, the samples were loaded into silica capillaries and temperature-induced fluorescence changes were measured on a Monolith™ (NanoTemper) at 25 °C by using 20% LED-power and 40% MST-power. Data analyses were performed by using the NTA software (NanoTemper Technologies). The experiments were repeated three times using different protein preparations.

DNA shape analysis and electrostatic potential calculation

Bound DNA conformation in the crystal structure was analyzed with CURVES+ ver.1.31³⁰. All theoretical width and electrostatic potential of the minor groove for DNA motifs in the *CUC1*, *CUC2*, and *NAC004* were calculated using the DNASHape server³¹.

Circular dichroism analysis

The circular dichroism (CD) signals of the samples were measured on the 4B8 beamline at the Beijing synchrotron radiation facility. Spectra were collected at 1 nm intervals over the wavelength ranging from 260 to 175 nm in a 0.005 cm optical path length at 25 °C. Protein samples were prepared at a concentration of about 10 mg mL⁻¹. A pure solvent baseline collected with the same cell was subtracted. All spectra data were processed using the CD

tool software package^{52,53}. The differential absorbance of polarized lights (mdeg) was converted into the per residue molar absorption unit, delta epsilon ($\Delta\epsilon$) in $M\text{ cm}^{-1}$, by normalization with respect to polypeptide concentration and path length.

Small-angle X-ray scattering

Small-angle X-ray scattering measurements were performed on beamline BL19U2 at the SSRF following previously published methods^{54,55}. Briefly, all proteins were subjected to size exclusion chromatography in a buffer containing 20 mM HEPES, pH 7.3, 500 mM NaCl, 2 mM β -mercaptoethanol. The REF6-DNA complex was dialyzed in a buffer containing 20 mM MES (pH 6.0), 150 mM NaCl, and 2 mM β -mercaptoethanol. Various concentrations of protein were used, and the data were collected at 1.03 Å with a distance of 2.68 m from the detector. Individual data were processed by RAW⁵⁶. The scattering data from the buffer alone were measured before and after each sample measurement, and the average of the scattering data was used for background subtraction. Data collection statistics are summarized in Supplementary Table S3⁵⁷.

Bimolecular fluorescence complementation assay

BiFC measurements were performed as described⁵⁸ using a Zeiss LSM710 confocal microscope at 514 nm. The full-length REF6 was cloned into the pSPYCE and pSPYNE BiFC vectors. The constructs were transfected into *A. tumefaciens* strain GV3101. The transfected bacteria were used to infiltrate the lower epidermis of tobacco leaves. After 72 h, the fluorescence signals were measured.

GST pull-down assay

REF6 constructs fused with His6-MBP or GST-tags were first purified using the appropriate affinity columns. Two hundred micrograms of His6-MBP-REF6 was incubated with GST-REF6 for 6 h at 4 °C in a 500 μ L buffer of 20 mM MES (pH 6.0), 150 mM NaCl, 2 mM β -ME and 0.1% (v/v) Triton X-100 in the presence of glutathione agarose beads (GE). The resin was extensively rinsed with the same buffer to remove unbound or nonspecifically bound proteins. Proteins left on the beads were separated by SDS-PAGE and analyzed by Western blotting.

Analytical ultracentrifugation

Sedimentation velocity (SV) analytical ultracentrifugation experiments were performed in a Beckman/Coulter XL-I analytical ultracentrifuge using double-sector or six-channel centerpieces and sapphire windows. The proteins were further purified by size exclusion chromatography in buffer containing 20 mM HEPES (pH 7.3) and 200 mM NaCl. The REF6-DNA complex was obtained by directly incubating the REF6 protein with an excess (1:3) of the

specific DNA substrate in 20 mM MES (pH 6.0), 150 mM NaCl, for 2 h on ice. The proteins were concentrated to 1 mg mL^{-1} before the experiments. SV experiments were conducted at 42,000 rpm and 4 °C using interference detection and double-sector cells loaded. The data were analyzed using the SEDPHAT and SEDFIT programs^{59,60}.

Chemical crosslinking

The fragments of REF6^{1223–1360} and REF6^{1239–1360} were purified using affinity columns. Fifteen micrograms of REF6^{1223–1360} and REF6^{1239–1360} was incubated with glutaraldehyde at different concentrations in a 100 μ L reaction volume. After incubation for 10 min at 25 °C, the crosslinking reaction was quenched with 1 M Tris, to produce a final concentration of 50 mM, and then incubated for 10 min at 25 °C. The samples were analyzed by SDS-PAGE followed by Coomassie Bright Blue staining⁶¹.

Acknowledgements

We would like to thank the staff of beamlines BL19U1, BL17U and BL19U2 at the Shanghai Synchrotron Radiation Facility and beamlines 3W1A and 4B8 at the Beijing Synchrotron Radiation Facility for the excellent technical assistance. We thank Prof. Dr. Xiaofeng Cao for generously providing REF6 cDNA. We thank Dawei Li and Zhaolei Li for the help with the BiFC assays and discussions. This work was supported financially by National Key Research and Development Program of China [2018YFE0113100]; National Natural Science Foundation of China [grant numbers 31872713, 31800629, 31570725, and 91519332]; State Key Laboratory of Agrobiotechnology Grant [2018SKLAB1-3]; Extramural Scientists of State Key Laboratory of Agrobiotechnology Grant [2019SKLAB6-17], and Hong Kong Research Grants Council Area of Excellence Scheme [AoE/M-403/16].

Author details

¹State Key Laboratory of Agrobiotechnology and Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Biological Sciences, China Agricultural University, 100193 Beijing, China. ²Department of Chemistry and Biochemistry, University of Windsor, Windsor, ON N9B 3P4, Canada. ³School of Life Sciences and Center for Soybean Research of the State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China. ⁴Department of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

Author contributions

Z.T. performed the gene cloning, protein expression, purification, crystallization, the diffraction data collection and structure determination. Z.T., X.L., M.L., W.W., M.Z., C.T., Z.L., Y.L., Zhenhang C., M.Y., L.M., C.C., Y.T., H.-M.L., S.D., and Zhongzhou C. contributed to the experimental design, data analyses, and discussion. Z.T. and Zhongzhou C. conceived the project and wrote the manuscript.

Data availability

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data bank under accession numbers 6A57 and 6A58.

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies the paper at (<https://doi.org/10.1038/s41421-020-0150-6>).

Received: 16 October 2019 Accepted: 3 February 2020

Published online: 31 March 2020

References

- Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7**, 715–727 (2006).
- Slama, P. Two-domain analysis of JmjN-JmjC and PHD-JmjC lysine demethylases: detecting an inter-domain evolutionary stress. *Proteins* **86**, 3–12 (2018).
- Zhang, Q. et al. Structure-function analysis reveals a novel mechanism for regulation of histone demethylase LSD2/AOF1/KDM1b. *Cell Res.* **23**, 225–241 (2013).
- Zhang, S. et al. C-terminal domains of a histone demethylase interact with a pair of transcription factors and mediate specific chromatin association. *Cell Discov.* **1**, 15003 (2015).
- Lu, F. et al. Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice. *J. Integr. Plant Biol.* **50**, 886–896 (2008).
- Lu, F., Cui, X., Zhang, S., Jenuwein, T. & Cao, X. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. *Nat. Genet.* **43**, 715–719 (2011).
- Crevillen, P. et al. Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature* **515**, 587–590 (2014).
- Zheng, S. et al. The Arabidopsis H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor. *Nat. Commun.* **10**, 1303 (2019).
- Noh, B. et al. Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of Arabidopsis flowering time. *Plant Cell* **16**, 2601–2613 (2004).
- Yu, X. et al. Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **105**, 7618–7623 (2008).
- Persikov, A. V. et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res.* **43**, 1965–1984 (2015).
- Yang, Y., Yin, X., Yang, H. & Xu, Y. Histone demethylase LSD2 acts as an E3 ubiquitin ligase and inhibits cancer cell growth through promoting proteasomal degradation of OGT. *Mol. Cell* **58**, 47–59 (2015).
- Cui, X. et al. REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in Arabidopsis. *Nat. Genet.* **48**, 694–699 (2016).
- Li, C. et al. Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in Arabidopsis. *Nat. Genet.* **48**, 687–693 (2016).
- Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
- Eom, K. S., Cheong, J. S. & Lee, S. J. Structural analyses of zinc finger domains for specific interactions with DNA. *J. Microbiol. Biotechnol.* **26**, 2019–2029 (2016).
- Hashimoto, H. et al. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.* **28**, 2304–2313 (2014).
- Hirano, Y. et al. Structure of the SHR-SCR heterodimer bound to the BIRD/IDD transcriptional factor JKD. *Nat. Plants* **3**, 17010 (2017).
- Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* **79**, 213–231 (2010).
- Nolte, R. T., Conlin, R. M., Harrison, S. C. & Brown, R. S. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. USA* **95**, 2938–2943 (1998).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
- He, F. et al. Solution structure of the zinc finger HIT domain in protein FON. *Protein Sci.* **16**, 1577–1587 (2007).
- Xu, R. X., Pawelczyk, T., Xia, T. H. & Brown, S. C. NMR structure of a protein kinase C-gamma phorbol-binding domain and study of protein-lipid micelle interactions. *Biochemistry* **36**, 10709–10717 (1997).
- Wang, D. et al. Role for first zinc finger of WT1 in DNA sequence specificity: Denys–Drash syndrome-associated WT1 mutant in ZF1 enhancer affinity for a subset of WT1 binding sites. *Nucleic Acids Res.* **46**, 3864–3877 (2018).
- Elrod-Erickson, M., Benson, T. E. & Pabo, C. O. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451–464 (1998).
- Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* **152**, 352–364 (2013).
- Aida, M., Ishida, T., Fukaki, H., Fujisawa, H. & Tasaka, M. Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *Plant Cell* **9**, 841–857 (1997).
- Takada, S., Hibara, K., Ishida, T. & Tasaka, M. The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development* **128**, 1127–1135 (2001).
- Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: curves+. *Nucleic Acids Res.* **37**, 5917–5929 (2009).
- Zhou, T. et al. DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* **41**, W56–W62 (2013).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Deng, Z. et al. Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nat. Commun.* **6**, 7642 (2015).
- Liu, Z. et al. Structural and functional insights into the human Borjeson-Forsman-Lehmann syndrome-associated protein PHF6. *J. Biol. Chem.* **289**, 10069–10083 (2014).
- Xu, C. et al. DNA sequence recognition of human CXXC domains and their structural determinants. *Structure* **26**, 85–95 e83 (2018).
- Chiu, T. P., Rao, S., Mann, R. S., Honig, B. & Rohs, R. Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein-DNA binding. *Nucleic Acids Res.* **45**, 12565–12576 (2017).
- Rohs, R. et al. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
- Rohs, R. et al. The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248–1253 (2009).
- Haran, T. E. & Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **42**, 41–81 (2009).
- Hou, X. et al. Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. *Nat. Commun.* **5**, 4601 (2014).
- Hyun, Y. et al. Multi-layered regulation of SPL15 and cooperation with SOC1 integrate endogenous flowering pathways at the Arabidopsis shoot meristem. *Dev. Cell* **37**, 254–266 (2016).
- Li, C. et al. Verification of DNA motifs in Arabidopsis using CRISPR/Cas9-mediated mutagenesis. *Plant Biotechnol. J.* **16**, 1446–1451 (2018).
- Smaczniak, C. et al. Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc. Natl. Acad. Sci. USA* **109**, 1560–1565 (2012).
- Yan, W. et al. Dynamic and spatial restriction of Polycomb activity by plant histone demethylases. *Nat. Plants* **4**, 681–689 (2018).
- Qiu, Q. et al. DNA methylation repels targeting of Arabidopsis REF6. *Nat. Commun.* **10**, 2063 (2019).
- Gao, Y. et al. Human apo-SRP72 and SRP68/72 complex structures reveal the molecular basis of protein translocation. *J. Mol. Cell Biol.* **9**, 220–230 (2017).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
- Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
- Alexander, N., Woetzel, N. & Meiler, J. bcl:Cluster: a method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *IEEE Int. Conf. Comput. Adv. Bio. Med. Sci.* **2011**, 13–18 (2011).

52. Lees, J. G., Smith, B. R., Wien, F., Miles, A. J. & Wallace, B. A. CDtool—an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal. Biochem.* **332**, 285–289 (2004).
53. Li, Q. Q. et al. Selective inhibition of cancer cells by enzyme-induced gain of function of phosphorylated melittin analogues. *Chem. Sci.* **8**, 7675–7681 (2017).
54. Hu, Y. et al. The amino-terminal structure of human fragile X mental retardation protein obtained using precipitant-immobilized imprinted polymers. *Nat. Commun.* **6**, 6634 (2015).
55. Zhang, X. et al. Complex structures of the abscisic acid receptor PYL3/RCAR13 reveal a unique regulatory mechanism. *Structure* **20**, 780–790 (2012).
56. Nielsen, S. S. et al. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J. Appl. Crystallogr.* **42**, 959–964 (2009).
57. Trehwella, J. et al. 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr. D Struct. Biol.* **73**, 710–728 (2017).
58. Walter, M. et al. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438 (2004).
59. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
60. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124 (2003).
61. Argo, A. S., Shi, C., Liu, F. & Goshe, M. B. Performing protein crosslinking using gas-phase cleavable chemical crosslinkers and liquid chromatography-tandem mass spectrometry. *Methods* **89**, 64–73 (2015).
62. Chang, M. H., Chen, A. P. & Romero, M. F. NBCE1A dimer assemble visualized by bimolecular fluorescence complementation. *Am. J. Physiol. Ren. Physiol.* **306**, F672–F680 (2014).