8388–8398 *Nucleic Acids Research, 2019, Vol. 47, No. 16* *Published online 21 June 2019*
doi: 10.1093/nar/gkz557

Structural basis of specific DNA binding by the transcription factor ZBTB24

Ren Ren[1,†], Swanand Hardikar[1,†], John R. Horton[1], Yue Lu[1], Yang Zeng[1,2], Anup K. Singh[1], Kevin Lin[1], Luis Della Coletta[1], Jianjun Shen[1,2], Celine Shuet Lin Kong[3], Hideharu Hashimoto[4], Xing Zhang[1], Taiping Chen[1,2,*] and Xiaodong Cheng[1,2,*]

[1]Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, [2]Program in Genetics and Epigenetics, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA, [3]Program in Cancer Biology, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA and [4]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

Received October 13, 2018; Revised June 07, 2019; Editorial Decision June 11, 2019; Accepted June 13, 2019

ABSTRACT

***ZBTB24*, encoding a protein of the ZBTB family of transcriptional regulators, is one of four known genes—the other three being *DNMT3B*, *CDCA7* and *HELLS*—that are mutated in immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome, a genetic disorder characterized by DNA hypomethylation and antibody deficiency. The molecular mechanisms by which ZBTB24 regulates gene expression and the biological functions of ZBTB24 are poorly understood. Here, we identified a 12-bp consensus sequence [CT(G/T)CCAGGACCT] occupied by ZBTB24 in the mouse genome. The sequence is present at multiple loci, including the *Cdca7* promoter region, and ZBTB24 binding is mostly associated with gene activation. Crystallography and DNA-binding data revealed that the last four of the eight zinc fingers (ZFs) (i.e. ZF5-8) in ZBTB24 confer specificity of DNA binding. Two ICF missense mutations have been identified in the ZBTB24 ZF domain, which alter zinc-binding cysteine residues. We demonstrated that the corresponding C382Y and C407G mutations in mouse ZBTB24 abolish specific DNA binding and fail to induce *Cdca7* expression. Our analyses indicate and suggest a structural basis for the sequence specific recognition by a transcription factor centrally important for the pathogenesis of ICF syndrome.**

INTRODUCTION

ZBTB (zinc-finger- and BTB domain-containing) proteins, characterized by the presence of an N-terminal BTB (broad-complex, tram-track, and bric-à-brac) domain [also known as POZ (poxvirus and zinc finger) domain] and a C-terminal domain with tandem C2H2 Krüppel-type zinc fingers (ZFs), are a large family of evolutionarily conserved transcriptional regulators (1). The BTB domain mediates homomeric or heteromeric dimerization and interacts with histone deacetylase (HDAC)-containing corepressor complexes, such as the NCOR (nuclear receptor corepressor), SMRT (silencing mediator of retinoic acid and thyroid hormone receptor), BCOR (BCL6 corepressor) and SIN3A/3B complexes (2). The ZF domain mediates sequence-specific DNA binding and, in some cases, also mediates protein–protein interactions. For example, ZBTB16/PLZF (promyelocytic leukemia zinc finger) and ZBTB17/MIZ-1 (Myc-interacting zinc finger protein-1) reportedly interact with p300 and MYC, respectively, via their ZF domains (3,4). Generally considered transcriptional repressors, ZBTB proteins have been implicated in a variety of developmental and cellular processes. Of particular note, several ZBTB proteins, including ZBTB7A/LRF (lymphoma/leukemia-related factor), ZBTB16/PLZF, ZBTB27/BCL6 (B-cell lymphoma 6) and ZBTB32/PLZP (PLZF-like zinc finger protein), play regulatory roles in lymphoid development, function and malignancies (5–7).

ZBTB24 is one of several genes that are mutated in immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome, a rare autosomal recessive disorder (8,9). Cases carrying *ZBTB24* mutations are known as ICF type 2 or ICF2 (10). The other genes associated with

*To whom correspondence should be addressed. Tel: +1 713 834 6274; Email: xcheng5@mdanderson.org
Correspondence may also be addressed to Taiping Chen. Tel: +1 512 237 9479; Email: tchen2@mdanderson.org
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

ICF syndrome are *DNMT3B* (DNA methyltransferase 3B, ICF1), *CDCA7* (cell division cycle associated 7, ICF3) and *HELLS* (helicase, lymphoid-specific, ICF4) (11–14). Patients with ICF syndrome exhibit antibody deficiency, despite the presence of B and T lymphocytes, and suffer from recurrent infections. Other clinical features include facial dysmorphism, failure to thrive and mental retardation (15,16). At the cellular level, ICF syndrome is characterized by the loss of DNA methylation in special genomic regions, most notably satellite repeats in pericentric regions (17,18), which lead to juxtacentromeric heterochromatin decondensation and chromosomal rearrangements in lymphocytes (15,16,19).

The ZF domain of ZBTB24 contains eight tandem ZFs. While most mutations identified in ICF2 patients are nonsense or frameshift mutations, at least two missense mutations have been identified in the ZF domain that alters conserved zinc-binding cysteine residues (C383Y and C408G) (10,20), highlighting the functional importance of the ZF domain. How ZBTB24 modulates DNA methylation and antibody production is poorly understood. Recent work reveals that ZBTB24 positively regulates *CDCA7* expression (21,22), suggesting that CDCA7 acts as a downstream effector in mediating at least some of the biological functions of ZBTB24. However, the mechanisms by which ZBTB24 controls *CDCA7* expression, as well as other target genes, remain to be determined. In this study, we show that ZBTB24 specifically recognizes a 12-bp consensus DNA sequence, which is present at multiple loci in the genome, including a region in the core promoter of human and mouse *Cdca7*. The last four ZFs of ZBTB24 (i.e. ZF5-8) confer DNA-binding specificity. Although ZF1–4 are not required in binding the specific sequence motif *in vitro*, cellular and molecular assays demonstrate that both ICF2 missense mutations, C383Y in ZF4 and C408G in ZF5 (equivalent to C382Y and C407G in mouse ZBTB24), eliminate the abilities of ZBTB24 to bind its target DNA and induce *Cdca7* expression. To understand the effects of DNA sequence on the binding of this important disease-associated transcription factor, we determined the structures of ZF4-8 in complex with DNA.

MATERIALS AND METHODS

Plasmid vectors

The HA-tagged mouse ZBTB24 construct was generated by cloning the *Zbtb24* cDNA (GenBank: AAH55367.1) into the *pCAG-HA-IRESblast* vector (23). The ZBTB24 point mutations, C382Y and C407G, were introduced by PCR-based mutagenesis and confirmed by sequencing. The fragments of ZBTB24 comprising of tandem ZF1-8 (residues 288–519) with or without the C382Y or C407G mutation, ZF4-8 (residues 375–519) and its mutant R390S, and ZF5-8 (residues 403–519) were cloned into the *pGEX-6P-1* vector with a GST fusion tag. The *Cdca7* luciferase (*Cdca7-Luc*) reporter construct was generated by cloning a ∼1.6-kb *Cdca7* promoter region into the *pGL3-Basic* vector (Promega). Deletion of the ZBTB24-binding motif in *Cdca7-Luc*, resulting in *Cdca7Δ-Luc*, was generated by PCR-based mutagenesis. All primers used in this study are listed in Supplementary Table S1.

Mouse embryonic stem cell (mESC) culture and generation of stable clones

J1 mESC culture and generation of stable clones expressing HA-tagged wild-type (WT) or mutant ZBTB24 were performed as described previously (23–26). *Zbtb24*- or *Cdca7*-deficient mESC lines were generated by CRISPR or homologous recombination.

Western blot

Western blot was carried out as described previously (25,27). All antibodies used in this study are listed in Supplementary Table S2.

Luciferase reporter assay

Cdca7-Luc or *Cdca7Δ-Luc* was co-transfected with *pRL-TK* (internal control) and HA-tagged WT or mutant ZBTB24 in *Zbtb24*-deficient mESCs. The cell lysates were measured for luciferase activities using the dual-luciferase reporter assay system (Promega).

ChIP-Seq analysis

To identify genome-wide ZBTB24-binding sites, we performed ChIP-Seq analysis using two stable clones (HA#3 and HA#4) expressing HA-ZBTB24 established in *Zbtb24*-deficient mESCs. For each sample, ∼5 million mESCs were used for ChIP with HA antibody. ChIP-Seq libraries were constructed using a modified KAPA LTP Library Preparation Kit Illumina Platform (Kapa Biosystems) and sequenced in a 36-bp single read run on Illumina HiSeq 2500 instrument (Illumina).

Approximately 22–26 million reads were generated per sample. Sequenced DNA reads were mapped to mouse genome (mm10) using Illumina analysis pipeline CASAVA (version 1.8.2) and only the reads that were mapped to unique position were retained. About 70–94% reads were mapped to mouse genome, with 44–59% uniquely mapped. To avoid PCR bias, for multiple reads that were mapped to the same genomic position, only one copy was retained for further analysis. Approximately 9–13 million reads were finally used in peak calling and downstream analyses. Each read was extended by 150 bp to its 3′ end. The number of reads on each genomic position was rescaled to normalize the total number of mapped reads to 10M and averaged over every 10-bp window. The normalized values were displayed in UCSC Genome Browser. HOMER (version 4.6) was used for the initial peak calling with the peak size of 150 bp (28). Two parameter settings were used: (i) the stringent setting is the default setting of HOMER: fold change to input ≥4, *P*-value to input ≤ 1e-4, fold change to local signal ≥4 and *P*-value to local signal ≤ 1e-4 and (ii) a less stringent setting is fold change to input ≥2, *P*-value to input ≤ 1e-3, fold change to local signal ≥2 and *P*-value to local signal ≤ 1e-3. Then, the peaks of the two replicates called at the stringent setting were merged (allowing at most 200 bp) and only the merged peaks that overlapped the peaks of both replicates at the less stringent setting were called as the peaks of ZBTB24.

Each peak was assigned to the gene that has the closest transcription start site (TSS) to it. Then, the peak was classified by its location to the gene: upstream (−50 kb to −5 kb from TSS), promoter (−5 kb to +0.5 kb from TSS), exon, intron, TES (−0.5 kb to +5 kb from transcription end site) and downstream (+5 kb to +50 kb from transcription end site). The genes used to annotate the peaks are RefSeq genes downloaded from UCSC Genome Browser on March 6, 2013. HOMER was used to search motifs within ±100 bp of peak summits and to identify the existence of the motif from ±250 bp of peaks summits. CentriMo (29) from MEME Suite (version 4.10.2) was used to evaluate how the motif was centrally enriched in ZBTB24 peaks (±500 bp from the summits). All the parameters were set as default. To assess the distribution of ChIP-Seq signal, each read was extended by 150 bp to its 3′ end. For each ZBTB24 peak, ±2 kb around the peak center was subdivided into 25-bp bins and the average read coverage per million total reads in each bin was calculated. The calculated values of ZBTB24 were then subtracted by the values of input, and averaged over the peaks with or without ZBTB24 motif.

For the heat maps and *K*-means clustering of ZBTB24 and chromatin marks, the raw fastq files of chromatin marks in Bruce4 mESCs were downloaded from ENCODE (30) and were processed in the same way as our ZBTB24 ChIP samples except that Bowtie (Version 1.1.2) (31) allowing two mismatches were used for mapping. For each mark, if it had replicates, the replicates were merged. From each ZBTB24 peak center, ±10 kb range was subdivided into 200-bp bins. For ZBTB24 and each chromatin mark, the rpkm (reads per kilobase per million mapped reads) value for each bin was calculated and subtracted by the rpkm of the corresponding input. The resulting value tables for ZBTB24 and all chromatin marks were put together and plotted in heat map by R function heatmap.2. *K*-means clustering based on the values of all chromatin marks was performed to group ZBTB24 peaks.

Protein expression and purification

Escherichia coli strain BL21 (DE3) Codon-plus was transformed with plasmid vectors expressing GST-ZBTB24 fragments comprising of ZF1-8, C382Y or C407G mutant in the context of ZF1-8, ZF4-8 (pXC1444), ZF5-8 (pXC1443) and R390S mutant in the context of ZF4-8 (pXC2071). Bacteria were grown in LB broth at 37°C until reaching the log phase ($A_{600\,nm}$ between 0.4 and 0.5), when the shake temperature was set to 16°C and 25 μM $ZnCl_2$ was added to the cell culture. When the shake temperature reached 16°C and $A_{600\,nm}$ reached ∼0.8, the protein expression was induced by the addition of 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) with subsequent growth for 20 h at 16°C. Cell harvesting and protein purification were carried out at 4°C through a four-column chromatography protocol, conducted in a BIO-RAD NGC™ system. Cells were collected by centrifugation and the pellet was suspended in lysis buffer [20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) and 25 μM $ZnCl_2$]. Cells were lysed by sonication and 0.3% (w/v) polyethylenimine was slowly titrated into the cell lysate before centrifugation (32). The debris was removed by centrifugation for 30 min at 47 000 *g* and the supernatant was loaded onto a 5 ml GSTrap column (GE Healthcare). The resin was washed by lysis buffer and the bound protein was eluted in 100 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP and 20 mM reduced form glutathione. GST-ZF4-8 and GST-ZF5-8 were digested with PreScission protease (produced in-house) to remove the GST fusion tag. The cleaved protein was loaded onto a 5 ml SP column (GE Healthcare). The protein was eluted by NaCl gradient from 0.25 to 1 M in 20 mM Tris-HCl, pH 7.5, 5% glycerol and 0.5 mM TCEP. The peak fractions were pooled, concentrated and loaded onto a HiLoad 16/60 Superdex S200 column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.5, 250 mM NaCl, 5% glycerol and 0.5 mM TCEP. GST-ZF1-8 and the C382Y and C407G mutants were used as GST fusion proteins.

Protein crystallization and structure determination

The protein–DNA complex was prepared by mixing 1 mM ZF4-8 fragment and various double-stranded DNA oligo (in 10 mM Tris-HCl, pH 7.5, and 100 mM NaCl) with molar ratio 1:1.2 with overnight incubation on ice. The protein–DNA complex crystals were grown using the sitting drop vapor diffusion method at 19°C under two conditions of 20–25% (w/v) polyethylene glycol 3350 with low pH (0.1 M BIS-TRIS, pH 5.5) or higher pH (0.1 M Hepes, pH 7.5) in the presence of 0.2 M ammonium acetate (or sodium chloride or potassium chloride or sodium malonate or potassium thiocyanate).

Crystals were flash frozen using 20% (v/v) ethylene glycol as the cryo-protector. The X-ray diffraction data were collected at SER-CAT 22-ID beamline of the Advanced Photon Source (Argonne National Laboratory) and processed by HKL2000 (33) or at our local Rigaku facility equipped with a MicroMax-003 Microfocus sealed tube X-ray generator, an AFC11 partial-χ, 4-axis goniometer and an HyPix-6000HE hybrid photon counting detector (Supplementary Table S3). The datasets collected from home source were processed with CrysAlis^{Pro} (Rigaku) and were scaled and merged with the AIMLESS program of the CCP4 interface (34). Molecular replacement was performed with the PHENIX PHASER module (35) first by using three ZFs from a structure of PRDM9 (PDB ID: 5V3G) identified by the PHYRE2 server (36) and an 18-bp DNA (PDB ID: 5MHK) as search models. Zn atom positions were identified by the anomalous peaks utilizing the processed dataset keeping Bijvoet pairs separated. PHENIX REFINE (37) was used for refinement with 5% randomly chosen reflections for validation by the *R*-free value. COOT (38) was used for model building and corrections between refinement rounds. Structure quality was analyzed during PHENIX refinements and later validated by the PDB validation server. Molecular graphics were generated using PyMol (Schrödinger, LLC).

Fluorescence-based DNA binding assay

Fluorescence polarization method was used to measure the binding affinity with a Synergy 4 Microplate Reader (BioTek). FAM-labeled DNA duplex (5 nM) was incubated

with varied amount of protein in 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% glycerol and 0.5 mM TCEP for 30 min at room temperature before measurement. The data were processed using Graph-pad prim (version 7.0) with equation [mP] = [maximum mP] × [*C*] / (K_D + [*C*]) + [baseline mP], in which mP is millipolarization and [*C*] is protein concentration. The K_D value for each protein–DNA interaction was derived from two replicated experiments.

Electrophoretic mobility shift assay (EMSA)

EMSA was performed on 8% native 1× TBE polyacrylamide gel. The individual sample was prepared by incubating 5 nM FAM-labeled DNA probe (FAM-5′-TCCACTGCCAGGACCTTT-3′/3′-GGTGACGGTCCTGGAAAA-5′) with variant amount of proteins in assay buffer (20 mM Tris-HCl, pH 7.5, 225 mM NaCl, 5% glycerol, 0.5 mM TCEP and 0.1 mg/ml BSA) for 1 h at room temperature, the samples were then loaded into each lane of native gel and ran for 55 min in 0.5× TBE buffer (100 V). Gel was imaged using 9410 Typhoon variable mode imager (GE Healthcare).

RESULTS

Mouse and human ZBTB24 bind a similar DNA sequence motif

Mouse ZBTB24, consisting of 710 amino acids (Figure 1A), and human ZBTB24, consisting of 697 amino acids, are highly conserved, with their BTB and ZF domains showing ∼96% and ∼95% sequence identity, respectively. In addition to the BTB and ZF domains, ZBTB24 also contains an AT-hook (Figure 1A), a short sequence motif that mediates interactions with the minor groove of AT-rich DNA as in HMGA1 (39). To determine the DNA-binding specificity of mouse ZBTB24, we performed ChIP-Seq with HA antibody using *Zbtb24*-deficient mESCs reconstituted with HA-tagged mouse ZBTB24 with expression levels similar to that of endogenous ZBTB24 in WT mESCs (Supplementary Figure S1). Bioinformatics analysis of the ChIP-Seq data identified 420 ZBTB24-binding peaks, located mainly in promoter regions (33%), gene bodies (introns and exons, 24%) and upstream regions (24%) (Figure 1B). Heat map and *K*-means clustering analysis revealed that many ZBTB24 target loci are enriched with active promoter marks (H3K4me3, H3K9ac, POLR2A) and/or enhancer marks (H3K4me1, H3K27ac, EP300) in mESCs (Supplementary Figure S2), suggesting that ZBTB24 binding is mostly associated with gene activation. Motif discovery by HOMER identified a 12-bp consensus sequence [CT(G/T)CCAGGACCT] (*P*-value = 1e-267) (Figure 1C), which is highly similar to the ZBTB24-binding motif recently identified in human cells (22). The consensus sequence was enriched in the center of the ZBTB24-binding peaks (Figure 1D). Among the 420 peaks, 209 (∼50%) were found to have the motif. In general, the peaks with the motif tended to have stronger signal than those without the motif (Figure 1E). We did not find the ZBTB24 recognition motif to be similar to any existing motif in the databases. Notably, a distinct peak was observed in the core promoter of *Cdca7* at a putative ZBTB24 recognition motif (CTGgCAGGACCT), located at -93 to -82 relative to TSS (Figure 1F), confirming that *Cdca7* is a direct target of ZBTB24. The ChIP-Seq genome browser screenshots of several other ZBTB24 direct target genes with the motif (*Taf6*, *Cdc40* and *Ostc*) are shown in Supplementary Figure S3. Consistent with the notion that ZBTB24 binding is generally associated with gene activation, RT-qPCR analysis revealed that *Cdca7*, *Taf6*, *Cdc40* and *Ostc* were all downregulated in *Zbtb24*-deficient mESCs (Supplementary Figure S4).

ZBTB24 ZF5–8 fingers are sufficient for specific DNA binding

To gain insights into the molecular mechanism by which the ZBTB24 tandem ZF array recognizes its target DNA sequence, we first compared the experimentally determined ChIP-Seq motif, identified in this study (Figure 1C) and also by Thompson *et al.* in human cells (22), to the predicted DNA-binding specificity of ZBTB24 by a computational algorithm (40) (Figure 2A). The predicted sequence for the last two fingers (ZF7-8) matches particularly well with the consensus, while the sequence responsible by ZF5-6 varies. We reasoned that the last four fingers (ZF5–8) are sufficient to bind the 12–13 bp consensus, because in conventional C2H2 ZF proteins, each finger interacts with three adjacent DNA base pairs (41,42). We generated a construct of ZBTB24 that includes the last five fingers (ZF4–8) and analyzed its binding to the double-stranded oligonucleotides containing the consensus sequence and an arbitrary negative control that partially overlaps the consensus. Fluorescence polarization was used to measure the dissociation constants (K_D) between the ZF protein and oligos (32). ZF4–8 bound the consensus oligo with a K_D of ∼25 nM and displayed >80 fold greater binding against the negative control that shares 5 bp within the consensus (Figure 2B). EMSA showed that deletion of ZF4 does not affect specific DNA binding (Figure 2C). These results confirm the presumed specificity of ZBTB24.

Structural investigations

We crystallized ZF4-8 bound with 19-bp oligos containing either the consensus sequence (TG=CTGCCAGGACCTG or TT=CTTCCAGGACCTG) or an altered sequence at nucleotide positions 2 and 3 (GT=CGTCCAGGACCTG). To facilitate the crystallization, we used a 5′-overhanging thymine on the top strand and a 5′-overhanging adenine on the bottom strand. The crystals of the protein–DNA complex was formed in space group $P2_1$, containing one complex per crystallographic asymmetric unit, with the DNA molecules coaxially stacked, with the overhanging A and T forming a base pair with neighboring DNA molecules, thus forming a pseudo continuous duplex. We determined a total of six structures to the resolutions of 1.54–1.99 Å (Supplementary Table S3). The six structures differ in the DNA sequences corresponding to the ZF4 and ZF5 binding, and we will first describe the common features and discuss the differences among them.

ZF4 spans along the DNA phosphate backbone and the last four fingers (ZF5–8) interact with DNA exclusively in

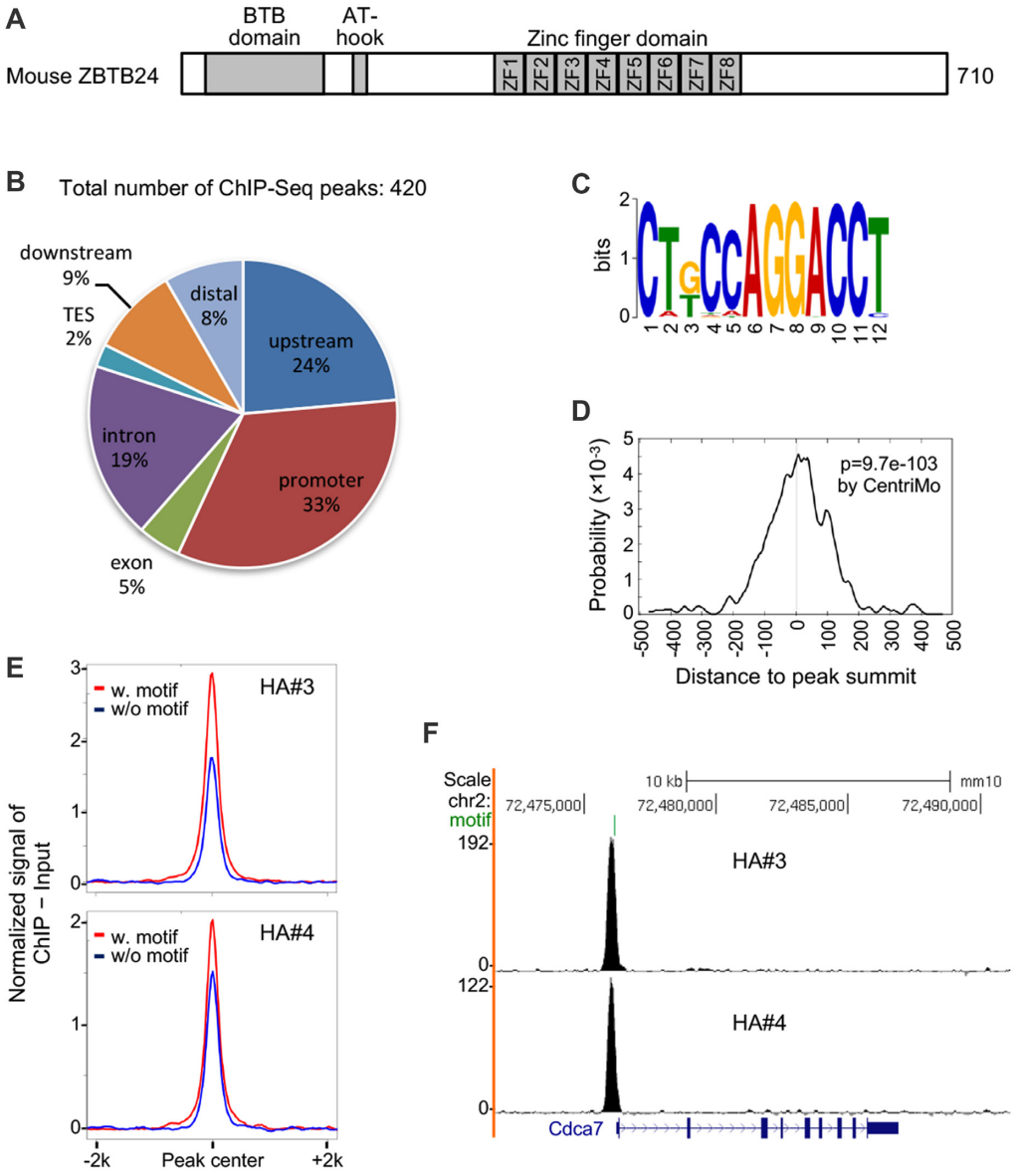

Figure 1. ChIP-Seq analysis identifies a 12-bp consensus DNA sequence recognized by ZBTB24. (**A**) Mouse ZBTB24 protein is schematically shown, including the BTB domain, AT-hook, and zinc finger (ZF) domain with eight tandem ZFs. (**B**) Pie chart showing the distribution of HA-ZBTB24 ChIP-Seq peaks in the genome. Upstream, -50 kb to -5 kb from TSS; promoter, −5 kb to +0.5 kb from TSS; TES, −0.5 kb to +5 kb from TES; downstream, +5 kb to +50 kb from TES; distal, other regions. (**C**) ZBTB24-binding motif identified from ChIP-Seq data. (**D**) CentriMo analysis showing the distribution of the motif to the peak summits. (**E**) Distribution of ZBTB24 signal over ±2 kb around the centers of ZBTB24 peaks with (red) and without (blue) the ZBTB24 motif. The signal of input was subtracted from the signal of ZBTB24. HA#3 and HA#4 are the two HA-ZBTB24 stable clones used for ChIP-Seq. (**F**) Genome browser screenshots showing a prominent ZBTB24-binding peak in the *Cdca7* promoter.

the major groove of the 13-bp consensus motif (Figure 2D). Among them, the last three fingers (ZF6–8) follow the one-finger-three base rule, whereas ZF5 recognizes four bases. The convention that we used for numbering nucleotides is that the consensus sequence is numbered 1–13 from 5′ (left) to 3′ (right) as the 'top' strand (colored blue in Figure 2E), whereas ZF5 interacts with the first four base pairs ($C_1T_2T_3C_4$), ZF6 interacts with the following 3-base triplet ($C_5A_6G_7$), followed by ZF7 ($G_8A_9C_{10}$), and ends with ZF8 interaction the 3′ sequence ($C_{11}T_{12}G_{13}$) (Figure 2E). Within each finger, potential base-interacting residues are highlighted in Figure 2F.

ZF7-8

The sequence-specific interactions mediated by ZF7-8 follow the established patterns of base recognition with interactions concentrated on the bottom strand (Figure 2E). R500, N503 and A506 of ZF8 and D472, A475 and R478 of ZF7 form base-interacting hydrogen bonds (H-bonds) and van der Waals contacts with each DNA triplet (Figure 3). Among them, two guanines of C:G pairs at base pair positions 11 and 10 are recognized by two arginines, R500 of ZF8 and R478 of ZF7 (Figure 3A and D). The terminal Nη1 and Nη2 groups of arginine donate H-bonds to the

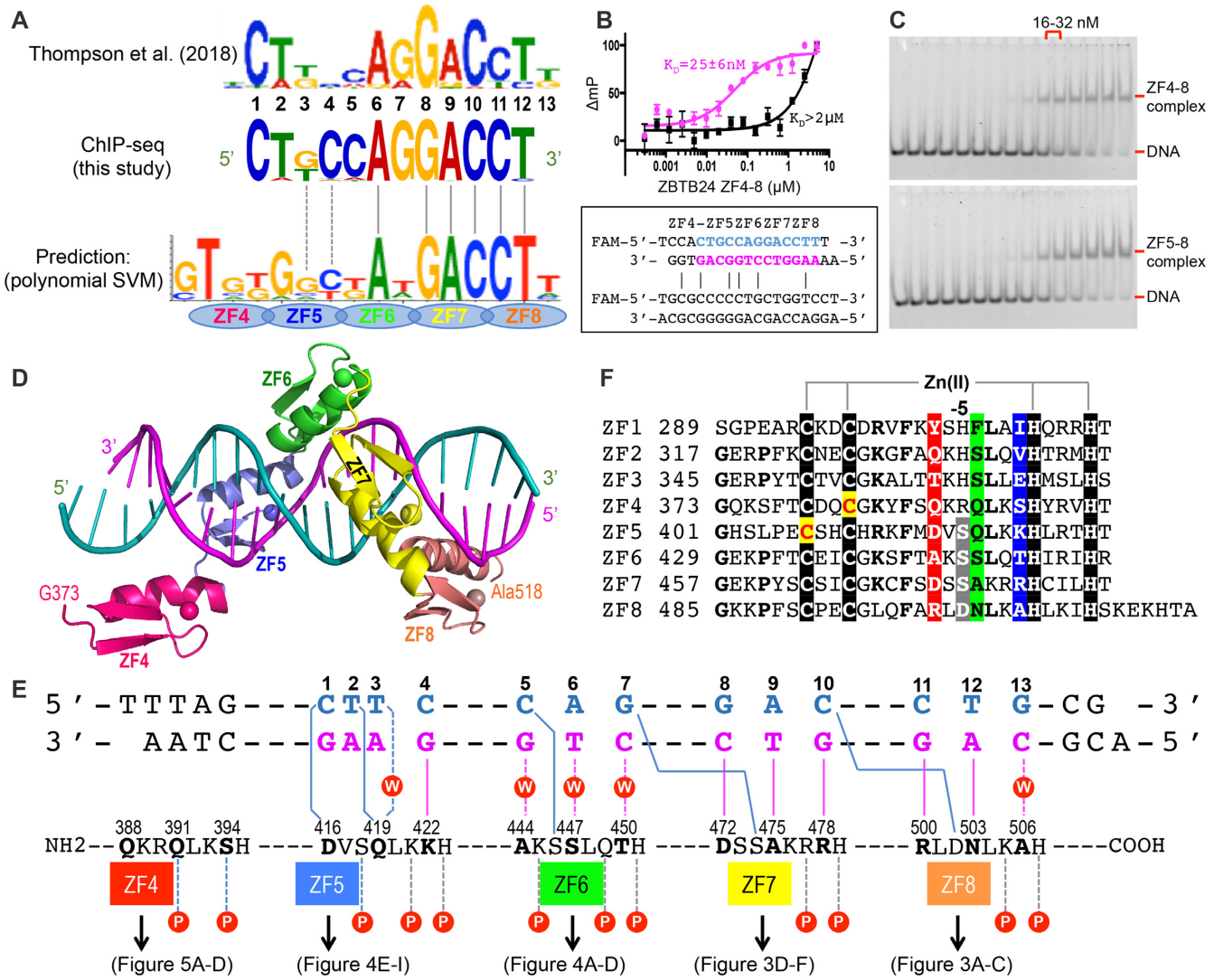

Figure 2. ZF4-8 is sufficient for binding the ZBTB24 consensus sequence. (**A**) ZBTB24 consensus binding motif as determined by ChIP-Seq in human (22) and mouse (this study). Included for comparison is the prediction made by polynomial SVM (support vector machines)-based algorithm (40). (**B**) The binding of ZF4-8 fragment against ZBTB24 consensus sequence. (**C**) EMSA of ZF4-8 (top) and ZF5-8 (bottom) binding to oligonucleotides containing the consensus sequence. The protein concentrations used were at maximum of 0.5 μM (the right most lane of each panel) followed by serial 2-fold dilutions (from right to left). (**D**) A structure of ZF4-8 (colored) binding ZBTB24 consensus sequence at the resolution of 1.59 Å (PDB ID: 6ML4). (**E**) General scheme of interactions between DNA and ZF4-8. The top two lines indicate the sequence of the double-stranded oligo used for crystallization. The base pair matching the consensus sequence are numbered as 1–13. The next line indicates amino acids of each finger (ZF4-8) from N to C terminus. (**F**) Sequence alignment of the eight zinc fingers of mouse ZBTB24. The four Zn-coordinating residues of each finger are highlighted with white letters against black (the two Cys residues mutated in ICF are colored in yellow). The amino acids colored in blue, green and red corresponds to the DNA base interactions (indicated in panel E). The residues colored gray at position -5 are in a sequence position before the first zinc-coordination His in each finger.

O6 and *N7* atoms of guanines, respectively. Similarly, the side chain of N503 forms bidentate contacts with adenine at base pair position 12, donating one H bond to adenine *N7* and accepting one from adenine *N6* (Figure 3B). D472 and A475 of ZF7 and A506 of ZF8, respectively, interact with three pyrimidines, cytosine at positions 8, thymine at position 9 and cytosine at position 13 (Figure 3C, E and F). D472 forms an H-bond with the cytosine-8 exocyclic amino group (Figure 3F). A475 forms a van der Waals contact with the thymine-9 methyl group (Figure 3E). In the ChIP-Seq sequence motif, base pair 13 is a variable position (Figure 2A). In the current structure, A506 interacts with cytosine-13 via a water-mediated contact (Figure 3C).

In the current sequence used for crystallization, base pairs 12 and 13 constitute a CpA/TpG dinucleotide. In the structure containing the methylated CpA site (PDB ID: 6ML6), the methylated cytosine-13 interacts similarly with A506 (via a water molecule) and forms an additional van der Waals contact with the side chain of N503, forming a 5mC-Asn-Ade triad (Figure 3G). In addition to the base interactions involving the bottom strand of each triplet, S474 of ZF7 and D502 of ZF8 interact with the top strand of the last base pair of previous triplet (Figure 2E). Thus, D502 of ZF8 accepts a H-bond from cytosine-10 in the preceding triplet (Figure 3D), similar to that of cytosine-8 and D472

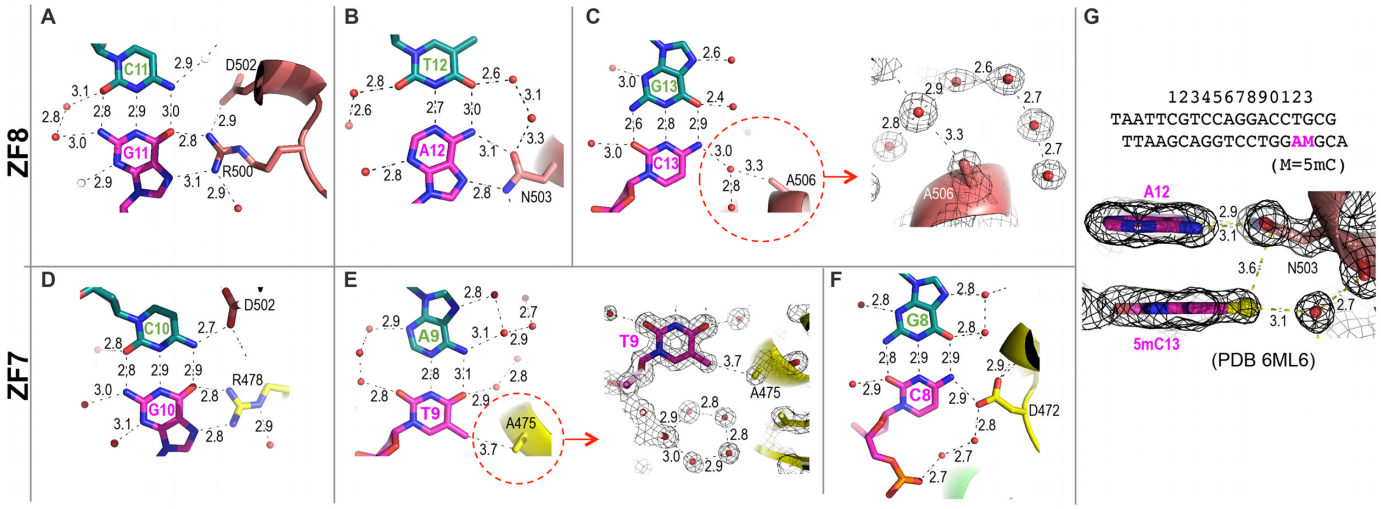

Figure 3. ZBTB24 ZF7-8 form base-specific contacts. (**A–C**) ZF8 interacts with a 3-bp triplet (base pairs 11–13). (A) R500 interacts with G_{11}. (B) N503 interacts with A_{12}. (C) A506 forms a water-mediated contact with C_{13}. (**D–F**) ZF7 interacts with a 3-bp triplet (base pairs 8–10). (D) R478 interacts with G_{10}. D502 of ZF8 interacts with the opposite C_{10}. (E) A475 contacts with T_9. A water ring formed next to the methyl group of T_9. (F) D472 interacts with C_8. (**G**) Methylated C_{13} forms a methyl-N503-Ade triad (PDB ID: 6ML6). Electron densities (2Fo–Fc) contoured at 1σ above the mean are shown in panels (C), (E) and (G).

interaction (Figure 3F), and S474 of ZF7 donates a H-bond to the *N7* atom of guanine-7 (see Figure 4D).

ZF6

Unlike ZF7-8, ZF6 contains small side chains at potential base-interacting residues (A444, S446, S447 and T450) (Figure 2E). S446 of ZF6 does accept a weak H-bond (3.2 Å) from the *N4* group of cytosine-5 (top strand) of the first base pair of the cognate triplet (Figure 4A). S447 forms a water-mediated interaction with *O4* atom of thymine-6 (Figure 4B). In addition, together with T443, S447 facilitates a layer of ordered water molecules enclosing the methyl group of thymine-6 (Figure 4C). In the ChIP-Seq motif, the paired adenine-6 is a highly conserved nucleotide (Figure 2A). Guanine-7 of top strand is contacted with S474 of ZF7 (the next finger) (Figure 4D). Together, three serine residues (S446, S447 and S474) interacting with three different pairs at base pair positions 5–7 seem to have adoptive interactions (Figure 4A–D), stemming in part from the ability of serine to act as an H-bond donor or acceptor or both at the same time.

ZF5

Unlike ZF6, ZF5 contains large and charged/polar residues (D416, Q419 and K422) at potential base-interacting positions (Figure 2E). Indeed, K422 of ZF5 interacts with guanine-4 (Figure 4E). We note that the base pair position 4 is a conserved cytosine in our ChIP-Seq motif, but a complete variable in the motif generated by Thompson *et al.* (Figure 2A). Unexpectedly, Q419 of ZF5 bridges between two base pairs, spanning two thymine bases of the top strand (Figure 4F; PDB ID: 6ML3) or between guanine-2 and adenine-3 from opposite strands (Figure 4G; PDB ID: 6ML4). The unusual one-residue spanning two-base pair recognition has recently been observed in ZFP568 (43), where a histidine spanning two base pairs. The expansion allowed the single unit of ZF4 to recognize 4-bp. In the ChIP-Seq sequence motif, base pairs 2 and 3, particularly position 3, are variable (Figure 2A). The next pair C:G at position 1 is recognized by D416, which accepts a H-bond from the *N4* group of cytosine-1 on the top strand (Figure 4H). The observed interactions for the three Asp-Cyt pairs are highly similar for Cyt1-D416 (ZF4), Cyt8-D472 (ZF7) and Cyt10-D502 (ZF8) (Figures 3D, F and 4H).

Glutamine can act as an H-bond donor or acceptor or both at the same time, explaining how Q419 might accommodate alternative base pairs. The variable sequence at base pairs 2 and 3 allowed us to introduce a CpG dinucleotide, and the methylation of the bottom strand cytosine-2 (in PDB ID: 6ML7) can be easily accommodated (Figure 4I).

ZF4

As expected from the size of consensus motif (12–13 bp), ZF4 is not involved in DNA sequence specific interactions, but instead it spans across the DNA major groove (Figure 2D), with limited interaction with DNA phosphate groups via Q391 and S394 (Figure 5A). Together with Q388, the three potential base-interacting residues are capable of making H-bonds and ZF4 folds as a normal finger in the structures we examined so far. We substituted DNA sequences (TTAG in PDB IDs: 6ML2 and 6ML3, GCTT in PDB ID: 6ML4, and AATT in PDB IDs: 6ML5, 6ML6 and 6ML7), prior to the consensus, assuming which would be recognized by ZF4, all resulted in the same ZF4 conformations (Figure 5B). It is interesting to note that the same set of protein residues (Q388, Q391 and S394) could potentially participate in DNA backbone phosphate interactions and DNA base specific H-bond interactions. We do not know

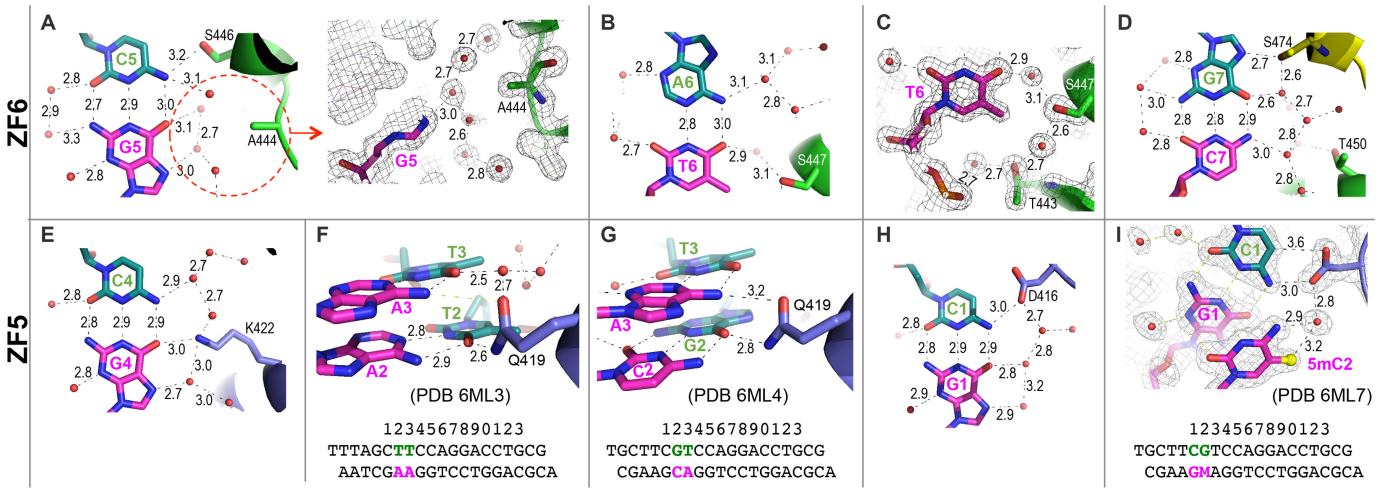

Figure 4. ZBTB24 ZF5-6 form base-specific contacts. (**A–D**) ZF6 interacts with a 3-bp triplet (base pairs 5–7). (A) S446 interacts with C_5. A444 is surrounded by a layer of hydration. (B) S447 conducts a water-mediated interaction with T_6. (C) A hydration shell enclosing the methyl group of T_6, connecting the *O4* atom to the phosphate oxygen of the same nucleotide. (D) S474 of ZF7 interacts with G_7. (**E–H**) ZF5 interacts with a 4-bp element (base pairs 1–4). (E) K422 interacts with G_4. (F) Q419 is capable of interacting with variable sequences at base pairs positions 2 and 3 (T_2T_3 in PDB ID: 6ML3). (G) Q419 spans between G_2 and A_3 (GT in PDB ID: 6ML4). (H) D416 interacts with C_1. (**I**) Methylation at C_2 is accommodated (PDB ID: 6ML7). Electron densities (2Fo–Fc) contoured at 1σ above the mean are shown in panels (A), (C) and (I).

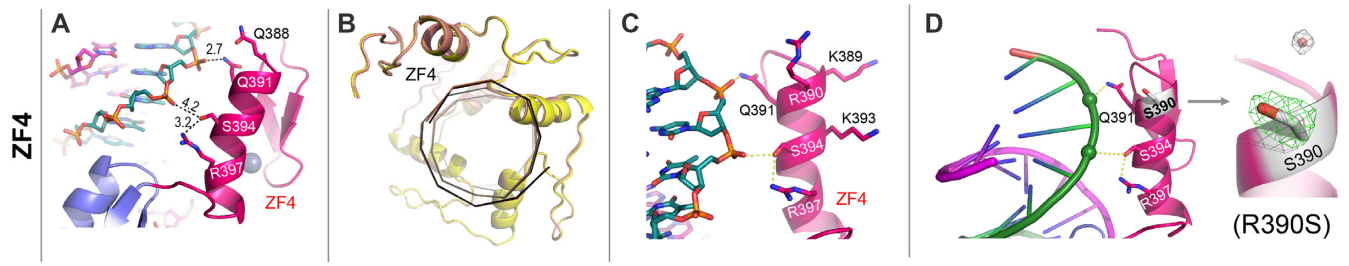

Figure 5. ZF4 is involved in non-specific DNA binding. (**A**) ZF4 interacts with DNA backbone phosphate groups via Q391 and S394. (**B**) Pairwise superimposition of structures with different sequences prior to the consensus motif, GCTT (PDB ID: 6ML4 in yellow) versus AATT (PDB ID: 6ML5 in brown). The room-mean-square deviation is <0.6 Å for 139 pairs of Cα atoms. (**C**) R390 is near but points away from DNA. (**D**) Mutant R390S adopts nearly identical conformation of ZF4 to that of wild-type. Omit electron density (Fo–Fc) contoured at 5σ above the mean is shown for the side chain of residue 390.

whether the intrinsic feature of ZF4 and/or its interfinger interactions prevent its binding in the DNA major groove.

We noticed that ZF1–4 fingers have a larger and positively charged residue (histidine or arginine) at position -5, whereas the corresponding residues in ZF5–8 are small (serine) or negatively charged (aspartate) (Figure 2F). The corresponding serine or aspartate residues in ZF6–8 interact with the top strand, the nucleotide immediately before or the first base pair of the cognate triplet (Figure 2E), as observed in PRDM9, a ZF protein containing 14 fingers all having a serine at the corresponding position (44). In the current structure (PDB ID: 6ML4), together with other positively charged residues (K389, K393, R397) from the same helix, R390 of ZF4 is near but pointing away from DNA backbone (Figure 5C). We asked whether the larger side chain of R390 of ZF4 could prevent ZF4 from binding in the DNA major groove. Thus, we substituted the arginine-to-serine (R390S). However, the structure of R390S mutant showed nearly identical conformation of ZF4 to that of wild-type (Figure 5D). Additional data will be required to settle the function of ZF4 and related ZF1-3.

Characterization of ICF2 missense mutations in the ZF domain of ZBTB24

As noted above, two ICF missense mutations have been identified in the ZF domain, i.e. C383Y and C408G in human ZBTB24 (10,20) (corresponding to C382Y of ZF4 and C407G of ZF5 in mouse ZBTB24; Figure 6A and B). Since ZF5, but not ZF4, is directly involved in DNA base specific interactions, we suspected that the C407G variant in ZF5 would abolish specific DNA binding and the C382Y variant in ZF4 would not. To test the idea, we produced GST fusion proteins of mouse ZBTB24 fragments comprising of ZF1–8 with no mutation (WT) or with the corresponding C382Y and C407G mutations (Figure 6C). To our surprise, EMSA revealed that both the C382Y and C407G mutants failed to bind a DNA probe containing the 12-bp ZBTB24 motif (Figure 6D). Given that ZF5–8 are sufficient for specific DNA binding (Figure 2C), our observation that the C382Y mutation in ZF4 abolished DNA binding suggests that the mutation has a neomorphic effect that interferes with ZF5-8-mediated DNA binding. There is an excellent precedent for the observation that an impaired ZF

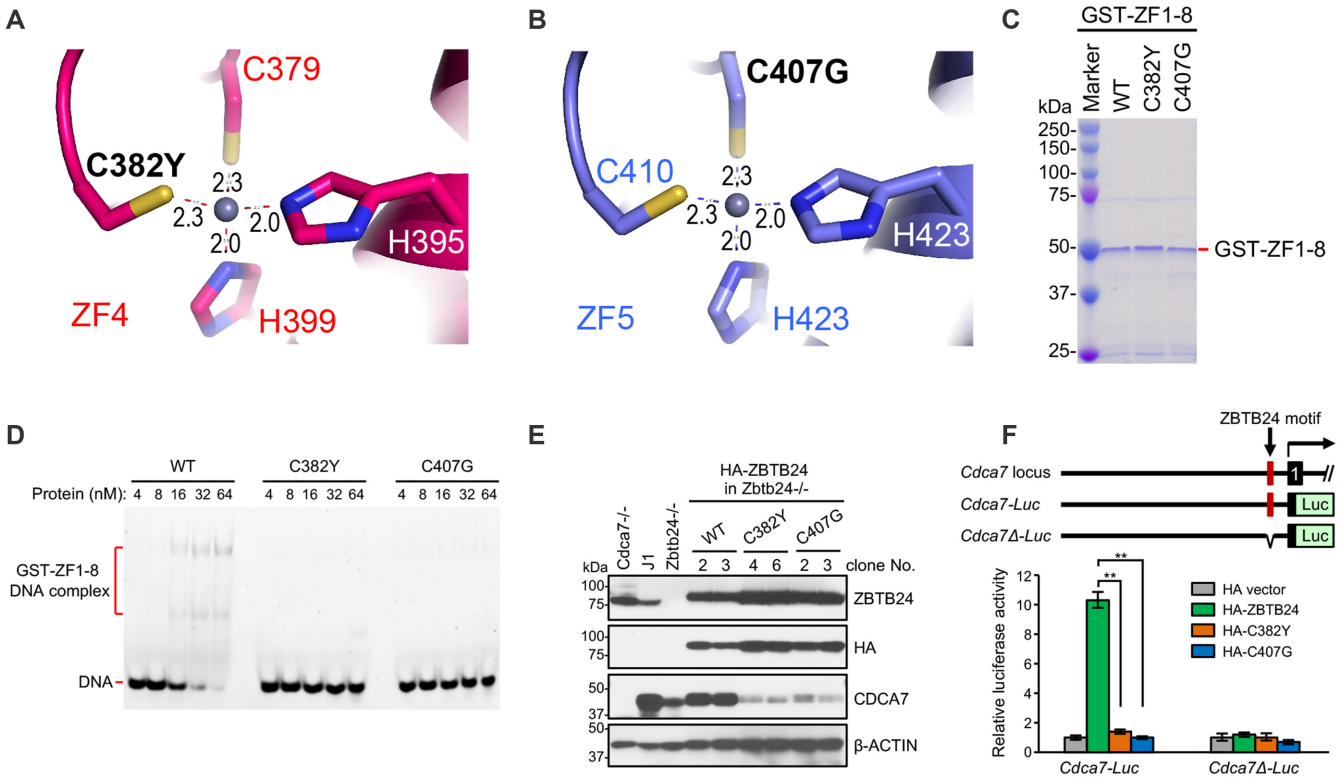

Figure 6. The C382Y and C407G mutations abolish the abilities of mouse ZBTB24 to bind specific DNA and induce CDCA7 expression. (**A** and **B**) C382 and C407 are respectively one of the four Zn ligands in ZF4 and ZF5. C382Y and C407G correspond to ICF missense mutations. (**C**) A Coomassie brilliant blue-stained gel showing purified GST-ZBTB24 fragments comprising of ZF1–8 with no mutation (WT) and with the C382Y and C407G mutations. (**D**) EMSA using GST fusion proteins described in panel (C) and a specific DNA probe containing the ZBTB24-binding motif (described in Figure 2C). The protein concentrations used are indicated. (**E**) Rescue experiments. HA-tagged WT or mutant ZBTB24 was stably expressed in *Zbtb24*-deficient mESCs and their ability to restore CDCA7 level was examined. Shown are western blots with the indicated antibodies. WT (J1), *Cdca7*$^{-/-}$ and *Zbtb24*$^{-/-}$ mESCs were used as controls. (**F**) Luciferase reporter assays. HA-tagged WT or mutant ZBTB24 or the empty HA vector was co-transfected with *Cdca7* promoter-driven luciferase reporter constructs in *Zbtb24*-deficient mESCs, and the cell lysates were used for luciferase activity assays 24 h post-transfection. *Cdca7-Luc*, reporter with intact ZBTB24-binding motif; *Cdca7Δ-Luc*, reporter with the ZBTB24-binding motif deleted. Luciferase activity results, relative to samples transfected with the empty HA vector, from three independent experiments (mean ± SD) are shown; $^{**}P < 0.01$.

distresses the DNA binding of neighboring fingers. An *Arabidopsis* histone demethylase REF6 contains four tandem ZFs with the first ZF containing only three zinc ligands (C2H1) (45)—a mimic of zinc-binding mutants. Whereas the isolated ZF2–4 of REF6 binds DNA, the addition of ZF1 might hijack a fourth zinc ligand (histidine) from the next finger. We note that the presence of similar histidine residues (H409 and H411) unique to ZF5 of ZBTB24 (Figure 3F) might allow the mutant ZF4 contact the ZF5 in a similar manner.

To assess the effects of the ICF2 missense mutations on CDCA7 expression, we stably expressed HA-tagged WT or mutant ZBTB24 in *Zbtb24*-deficient mESCs and performed rescue experiments. The CDCA7 level was largely recovered in *Zbtb24*-deficent mESCs reconstituted with WT ZBTB24 but showed no changes in cells reconstituted with the C382Y or C407G mutant protein (Figure 6E). To verify the importance of ZBTB24 binding to the *Cdca7* promoter in inducing transcription, we performed luciferase reporter assays in *Zbtb24*-deficient mESCs, using the *Cdca7* promoter region (~1.6 kb) with the 12-bp ZBTB24 binding motif being intact (*Cdca7-Luc*) or deleted (*Cdca7Δ-Luc*). Co-expression of HA-tagged WT ZBTB24, but not the C382Y or C407G mutant, led to a marked induction of the *Cdca7-Luc* reporter. This effect is dependent on the presence of the ZBTB24 binding motif, as HA-ZBTB24 failed to induce the *Cdca7Δ-Luc* reporter (Figure 6F).

DISCUSSION

Genetic defects in *ZBTB24* and *CDCA7* lead to similar disease phenotypes, including DNA hypomethylation and antibody deficiency, indicating that these genes function in the same pathway(s) in modulating developmental and cellular processes. Indeed, the ZBTB24-binding motif is conserved in the promoter regions of human and mouse *Cdca7* [the human (CTTGCAGGACCT) and mouse (CTGGCAGGACCT) sequences are identical except 1 bp (underlined) at position 3, which is a variable (G or T) in the motif identified by ChIP-Seq]. In this study, we verified the specificity of the interaction between ZBTB24 and its recognition sequence and validated the functional relevance of the interaction in inducing *Cdca7* expression with rescue experiments and luciferase reporter assays. It is worth mentioning that, among the 420 ZBTB24-binding peaks identified by ChIP-Seq, ~50% did not contain the sequence motif. As the BTB domain is known to interact with various proteins, including forming heteromeric dimers with other ZBTB proteins (2), some of the peaks without the motif

could be due to recruitment of ZBTB24 to specific genomic loci by interacting partners. Another possibility that cannot be ruled out is that ZF1–4 of ZBTB24 are also involved in DNA binding with weaker sequence preferences as compared to ZF5–8.

Most ZBTB proteins interact with HDAC-containing complexes via their BTB domains and function as transcriptional repressors (2). Our results, as well as the recent work by Thompson *et al.* (22), indicate that ZBTB24 binding is mostly associated with active chromatin marks. Indeed, examination of four ZBTB24 target genes (*Cdca7, Taf6, Cdc40* and *Ostc*) revealed that they were all downregulated in *Zbtb24*-deficient mESCs, confirming that ZBTB24 is a positive regulator. It would be interesting to determine whether the BTB domain of ZBTB24 recruits co-activator complexes. We compared our ZBTB24 ZF–DNA structures with previously characterized structures of ZBTB33 (also known as Kaiso) and ZBTB38 (46,47)—all are members of ZBTB family (Supplementary Figure S5). Among the three DNA-binding proteins, the DNA recognition sequences have limited similarity, reflecting the fact that the residues involved in base contacts are not conserved among the three proteins.

Two ICF missense mutations C382Y and C407G (equivalent to C383Y and C408G in human ZBTB24) are, respectively, one of the four zinc ligands in ZF4 and ZF5 (Figure 6A and B). We noted that mutations of Zn-coordination cysteine to tyrosine or glycine are common in another disease-associated ZF transcription factor, Wilms tumor protein [see Figure 1 of reference (48)]. The substitutions would disrupt the binding of zinc ion and thus fail to recognize the target DNA, thereby abolishing the ability to induce *Cdca7* promoter activity. Interestingly, ZF4 is not directly involved in DNA base specific interactions, but the C382Y mutant in ZF4 failed to bind specific DNA, suggesting a neomorphic effect that interferes with ZF5-8-mediated DNA binding.

In summary, we provide structural, biochemical and cellular data demonstrating that ZBTB24 activates *Cdca7* transcription by recognizing a specific DNA sequence in the *Cdca7* promoter. We also show that two ICF2 missense mutations identified in the ZF domain both eliminate the ability of ZBTB24 to bind its target DNA. Although ZF5–8 are sufficient for specific DNA binding, our results suggest that misfolding of ZF4, as a consequence of the C-to-Y mutation, prevents DNA binding mediated by ZF5–8.

DATA AVAILABILITY

The ChIP-Seq data have been deposited in the GEO database (accession number: GSE120969). The X-ray structures (coordinates and structure factors) of ZBTB24 ZF4–8 in complex with DNA (PDB IDs: 6ML2, 6LM3, 6ML4, 6ML5, 6ML6 and 6ML7) have been deposited to Protein Data Bank.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Z. Ying, B. Liu, H. Zhao, N. Veland, J. Dan and T. Hamidi in the Chen laboratory for discussion and technical assistance, B. Baker at New England Biolabs for synthesizing oligonucleotides, and A. Setili and Y. Cao in the Cheng laboratory for help with protein purification and technique supports.

Author Contributions: R.R. performed protein purification, DNA-binding experiments, crystallization and X-ray data collection; J.R.H. performed X-ray data collection and structure determination; S.H. generated *Zbtb24*- and *Cdca7*-deficient mESC lines and stable mESC clones expressing WT and mutant ZBTB24 and performed molecular and cellular experiments to characterize the ICF2 missense mutations; Y.Z. participated in molecular and cellular experiments; A.K.S. performed ChIP; L.D.C. performed deep sequencing under the supervision of J.S; Y.L. and K.L. performed bioinformatics analysis; H.H. made expression constructs and performed protein purification and DNA binding assays; C.S.L.K. and R.R. generated R390S mutant; X.Z. participated in experimental design and discussions throughout the study; T.C. and X.C. conceived the project, supervised the experimental work and wrote the manuscript.

FUNDING

U.S. National Institutes of Health (NIH) [1R01AI121403 0A1 to T.C.; GM049245-24 to X.C.; CA16672 to the CCSG Cores at MD Anderson Cancer Center]; the Cancer Prevention and Research Institute of Texas [RR160029 to X.C., RP170002 to J.S.]; Thomas Endowment fellowship (to Y.Z.). Funding for open access charge: NIH.

Conflict of interest statement. None declared.

This paper is linked to: https://doi.org/10.1093/nar/gkz381.

REFERENCES

1. Stogios,P.J., Downs,G.S., Jauhal,J.J., Nandra,S.K. and Prive,G.G. (2005) Sequence and structural analysis of BTB domain proteins. *Genome Biol.*, **6**, R82.
2. Perez-Torrado,R., Yamada,D. and Defossez,P.A. (2006) Born to bind: the BTB protein-protein interaction domain. *Bioessays*, **28**, 1194–1202.
3. Guidez,F., Howell,L., Isalan,M., Cebrat,M., Alani,R.M., Ivins,S., Hormaeche,I., McConnell,M.J., Pierce,S., Cole,P.A. *et al.* (2005) Histone acetyltransferase activity of p300 is required for transcriptional repression by the promyelocytic leukemia zinc finger protein. *Mol. Cell Biol.*, **25**, 5552–5566.
4. Moroy,T., Saba,I. and Kosan,C. (2011) The role of the transcription factor Miz-1 in lymphocyte development and lymphomagenesis-Binding Myc makes the difference. *Semin. Immunol.*, **23**, 379–387.
5. Lee,S.U. and Maeda,T. (2012) POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol. Rev.*, **247**, 107–119.
6. Siggs,O.M. and Beutler,B. (2012) The BTB-ZF transcription factors. *Cell Cycle*, **11**, 3358–3369.
7. Maeda,T. (2016) Regulation of hematopoietic development by ZBTB transcription factors. *Int. J. Hematol.*, **104**, 310–323.
8. Hulten,M. (1978) Selective somatic pairing and fragility at lq12 in a boy with common variable lmmuno deficiency. *Clin. Genet.*, **14**, 294.
9. Tiepolo,L., Maraschio,P., Gimelli,G., Cuoco,C., Gargani,G.F. and Romano,C. (1979) Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Hum. Genet.*, **51**, 127–137.
10. de Greef,J.C., Wang,J., Balog,J., den Dunnen,J.T., Frants,R.R., Straasheijm,K.R., Aytekin,C., van der Burg,M., Duprez,L., Ferster,A. *et al.* (2011) Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am. J. Hum. Genet.*, **88**, 796–804.

11. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
12. Xu,G.L., Bestor,T.H., Bourc'his,D., Hsieh,C.L., Tommerup,N., Bugge,M., Hulten,M., Qu,X., Russo,J.J. and Viegas-Pequignot,E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, **402**, 187–191.
13. Hansen,R.S., Wijmenga,C., Luo,P., Stanek,A.M., Canfield,T.K., Weemaes,C.M. and Gartler,S.M. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14412–14417.
14. Thijssen,P.E., Ito,Y., Grillo,G., Wang,J., Velasco,G., Nitta,H., Unoki,M., Yoshihara,M., Suyama,M., Sun,Y. *et al.* (2015) Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat. Commun.*, **6**, 7870.
15. Maraschio,P., Zuffardi,O., Dalla Fior,T. and Tiepolo,L. (1988) Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. *J. Med. Genet.*, **25**, 173–180.
16. Hagleitner,M.M., Lankester,A., Maraschio,P., Hulten,M., Fryns,J.P., Schuetz,C., Gimelli,G., Davies,E.G., Gennery,A., Belohradsky,B.H. *et al.* (2008) Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). *J. Med. Genet.*, **45**, 93–99.
17. Jeanpierre,M., Turleau,C., Aurias,A., Prieur,M., Ledeist,F., Fischer,A. and Viegas-Pequignot,E. (1993) An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum. Mol. Genet.*, **2**, 731–735.
18. Tuck-Muller,C.M., Narayan,A., Tsien,F., Smeets,D.F., Sawyer,J., Fiala,E.S., Sohn,O.S. and Ehrlich,M. (2000) DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. *Cytogenet. Cell Genet.*, **89**, 121–128.
19. Ehrlich,M., Sanchez,C., Shao,C., Nishiyama,R., Kehrl,J., Kuick,R., Kubota,T. and Hanash,S.M. (2008) ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity*, **41**, 253–271.
20. Nitta,H., Unoki,M., Ichiyanagi,K., Kosho,T., Shigemura,T., Takahashi,H., Velasco,G., Francastel,C., Picard,C., Kubota,T. *et al.* (2013) Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. *J. Hum. Genet.*, **58**, 455–460.
21. Wu,H., Thijssen,P.E., de Klerk,E., Vonk,K.K., Wang,J., den Hamer,B., Aytekin,C., van der Maarel,S.M. and Daxinger,L. (2016) Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. *Hum. Mol. Genet.*, **25**, 4041–4051.
22. Thompson,J.J., Kaur,R., Sosa,C.P., Lee,J.H., Kashiwagi,K., Zhou,D. and Robertson,K.D. (2018) ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Res.*, **46**, 10034–10051.
23. Kim,S.J., Zhao,H., Hardikar,S., Singh,A.K., Goodell,M.A. and Chen,T. (2013) A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood*, **122**, 4086–4089.
24. Chen,T., Ueda,Y., Dodge,J.E., Wang,Z. and Li,E. (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell Biol.*, **23**, 5594–5605.
25. Dan,J., Rousseau,P., Hardikar,S., Veland,N., Wong,J., Autexier,C. and Chen,T. (2017) Zscan4 inhibits maintenance DNA methylation to facilitate telomere elongation in mouse embryonic stem cells. *Cell Rep.*, **20**, 1936–1949.
26. Veland,N., Hardikar,S., Zhong,Y., Gayatri,S., Dan,J., Strahl,B.D., Rothbart,S.B., Bedford,M.T. and Chen,T. (2017) The arginine methyltransferase PRMT6 regulates DNA methylation and contributes to global DNA hypomethylation in cancer. *Cell Rep.*, **21**, 3390–3397.
27. Veland,N., Lu,Y., Hardikar,S., Gaddis,S., Zeng,Y., Liu,B., Estecio,M.R., Takata,Y., Lin,K., Tomida,M.W. *et al.* (2019) DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Res.*, **47**, 152–167.
28. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
29. Bailey,T.L. and Machanick,P. (2012) Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.*, **40**, e128.
30. Consortium,E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
31. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
32. Patel,A., Hashimoto,H., Zhang,X. and Cheng,X. (2016) Characterization of how DNA modifications affect DNA binding by C2H2 zinc finger proteins. *Methods Enzymol.*, **573**, 387–401.
33. Otwinowski,Z., Borek,D., Majewski,W. and Minor,W. (2003) Multiparametric scaling of diffraction intensities. *Acta Crystallogr. A*, **59**, 228–234.
34. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1204–1214.
35. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
36. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
37. 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.
38. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
39. Huth,J.R., Bewley,C.A., Nissen,M.S., Evans,J.N., Reeves,R., Gronenborn,A.M. and Clore,G.M. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.*, **4**, 657–665.
40. Persikov,A.V. and Singh,M. (2014) De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.*, **42**, 97–108.
41. Choo,Y. and Klug,A. (1997) Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.*, **7**, 117–125.
42. Wolfe,S.A., Nekludova,L. and Pabo,C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.
43. Patel,A., Yang,P., Tinkham,M., Pradhan,M., Sun,M.A., Wang,Y., Hoang,D., Wolf,G., Horton,J.R., Zhang,X. *et al.* (2018) DNA conformation induces adaptable binding by tandem zinc finger proteins. *Cell*, **173**, 221–233.
44. Patel,A., Zhang,X., Blumenthal,R.M. and Cheng,X. (2017) Structural basis of human PR/SET domain 9 (PRDM9) allele C-specific recognition of its cognate DNA sequence. *J. Biol. Chem.*, **292**, 15994–16002.
45. Qiu,Q., Mei,H., Deng,X., He,K., Wu,B., Yao,Q., Zhang,J., Lu,F., Ma,J. and Cao,X. (2019) DNA methylation repels targeting of Arabidopsis REF6. *Nat. Commun.*, **10**, 2063.
46. Nikolova,E.N., Stanfield,R.L., Dyson,H.J. and Wright,P.E. (2018) CH…O Hydrogen bonds mediate highly specific recognition of methylated CpG sites by the zinc finger protein kaiso. *Biochemistry*, **57**, 2109–2120.
47. Hudson,N.O., Whitby,F.G. and Buck-Koehntop,B.A. (2018) Structural insights into methylated DNA recognition by the C-terminal zinc fingers of the DNA reader protein ZBTB38. *J. Biol. Chem.*, **293**, 19835–19843.
48. Wang,D., Horton,J.R., Zheng,Y., Blumenthal,R.M., Zhang,X. and Cheng,X. (2018) Role for first zinc finger of WT1 in DNA sequence specificity: Denys-Drash syndrome-associated WT1 mutant in ZF1 enhances affinity for a subset of WT1 binding sites. *Nucleic Acids Res.*, **46**, 3864–3877.