3752–3764 *Nucleic Acids Research, 2019, Vol. 47, No. 7* *Published online 5 February 2019*
doi: 10.1093/nar/gkz077

Structural basis for DNA recognition by FOXC2

Xiaojuan Chen[1,2,†], Hudie Wei[1,†], Jun Li[1], Xujun Liang[1], Shuyan Dai[1], Longying Jiang[1], Ming Guo[1], Lingzhi Qu[1], Zhuchu Chen[1], Lin Chen[3] and Yongheng Chen[1,2,*]

[1]NHC Key Laboratory of Cancer Proteomics and Laboratory of Structural Biology, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China, [2]Key Laboratory of Medical Genetics and College of Life Science, Central South University, Changsha, Hunan 410008, China and [3]Molecular and Computational Biology Program, Department of Biological Sciences and Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA

Received October 03, 2018; Revised January 26, 2019; Editorial Decision January 28, 2019; Accepted January 30, 2019

ABSTRACT

The FOXC family of transcription factors (FOXC1 and FOXC2) plays essential roles in the regulation of embryonic, ocular, and cardiac development. Mutations and abnormal expression of FOXC proteins are implicated in genetic diseases as well as cancer. In this study, we determined two crystal structures of the DNA-binding domain (DBD) of human FOXC2 protein, in complex with different DNA sites. The FOXC2-DBD adopts the winged-helix fold with helix H3 contributing to all the base specific contacts, while the N-terminus, wing 1, and the C-terminus of FOXC2-DBD all make additional contacts with the phosphate groups of DNA. Our structural, biochemical, and bioinformatics analyses allow us to revise the previously proposed DNA recognition mechanism and provide a model of DNA binding for the FOXC proteins. In addition, our structural analysis and accompanying biochemical assays provide a molecular basis for understanding disease-causing mutations in FOXC1 and FOXC2.

INTRODUCTION

The Forkhead box (FOX) family of transcription factors, characterized by a conserved 'forkhead' or 'winged-helix' DNA-binding domain (DBD), regulate a wide variety of biological functions, including cell proliferation, differentiation, apoptosis, immunity, and metabolism (1,2). The FOX gene was originally identified in *Drosophila melanogaster*, and hundreds of members have been found subsequently in various species (3,4). So far, about 49 FOX genes have been identified in the human genome and categorized into 19 subgroups (FOXA to FOXS) (5).

The FOXC subgroup proteins, FOXC1 and FOXC2, share 90% full-length sequence identity, with 98% consensus in the forkhead domain (Supplementary Figure S1). FOXC1 is an essential component in embryonic development of brain (6), eye (7) and bone (8). FOXC1 knockout mice die at birth with hydrocephalus, eye defects, and multiple skeletal abnormalities (9). FOXC2, the other member of FOXC subfamily, is a key regulator of adipocyte metabolism (10), skeletal tissue development (11), lymphangiogenesis, (12) and lung maturation (13). In addition, FOXC1 and FOXC2 cooperatively control developmental processes such as cardiovascular, renal and somite development (14–16). FOXC1 is recently identified as a critical regulator for the niche formation of haematopoietic stem and progenitor cells (17), as well as hair follicle stem cells (18). FOXC2 is also required for maintaining murine spermatogonial stem cells (19). Furthermore, accumulating evidence identified emerging roles of FOXC proteins in initiation and development of cancers (20,21). Elevated expression of FOXC1 or FOXC2 occur in a variety of cancers such as breast carcinoma, hepatocellular carcinomas, lymphoma and so on (20,22). They promote cancers through mediating cell proliferation, metastasis, epithelial–mesenchymal transition (EMT), angiogenesis and lymphangiogenesis (23–26).

FOXC transcription factors play critical roles in many gene regulatory pathways, consistent with the fact that mutations of FOXC proteins result in developmental anomalies. A lot of mutations in FOXC1 and FOXC2 have been reported, including missense and nonsense mutations, small deletions and insertions. Mutations of FOXC1 are associated with Axenfeld–Rieger syndrome (ARS), an eye disease with autosomal dominant genetic changes primarily in anterior segment dysgenesis (27). About 31 missense FOXC1 mutants have been identified in ARS patients, with 29 mutations in its forkhead domain (22). Mutations in the DNA binding domain of FOXC2 have also been reported to cause Lymphoedema distichiasis syndrome (LDS), an inherited primary lymphedema (28).

*To whom correspondence should be addressed. Tel: +86 731 84327542; Fax: +86 731 84327542; Email: yonghenc@163.com
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Though all FOX proteins possess highly conserved DBDs, they have divergent DNA-binding properties and functions. Several three-dimensional structures of FOX-DBD proteins have revealed that the DNA binding domain is composed of three α-helices, three β-sheets and two less conserved winged loops (29–32). The main DNA recognition region is located in the third helix (H3). A previous solution NMR structure of FOXC2-DBD indicated its classic FOX DNA-binding domain, but the wing regions, which are divergent from other FOX proteins, are flexible without DNA interaction(33). In addition, the detailed mechanisms by which FOXC proteins bind DNA remain unknown. Hence, it is necessary to solve the structure of FOXC-DNA complex.

Here, we report the crystal structures of the FOXC2-DBD bound to two different DNA recognition elements at 2.40 and 2.32 Å respectively. Our structural, biochemical, and bioinformatics analyses not only provide the molecular basis of DNA recognition by FOXC proteins, but also help understand the mechanism by which disease-causing mutations affect protein function.

MATERIALS AND METHODS

Protein expression and purification

The DBD of human FOXC2 (amino acids 72–171) wild type or mutants were cloned into modified pET-28a vector, which contains a PreScission protease-cleavable N-terminal 6× His tag (34). Plasmids were confirmed by DNA sequencing (Genewiz, Suzhou, China) and then transformed into *Escherichia coli* Rosetta BL21 (DE3) cells and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 12 h at 25°C. Protein was purified by nickel affinity chromatography (GE healthcare) according to standard protocol. In order to crystallize, the His-tag of wild type was removed by incubation with PreScission protease at 4°C overnight. Then protein was further purified using cation-exchange chromatography (Mono S 5/50GL, GE healthcare) and size exclusion chromatography (superdex 75 10/300 GL, GE healthcare). Peak fractions were collected and concentrated to 28 mg/ml. The final protein was stored in 10 mM HEPES (pH 7.5), 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP at −80°C until use.

For isothermal titration calorimetry (ITC) measurements, the DBD of human FOXA2 (amino acids 163–264) was cloned to pGEX-6P1 vector and purified as described previously (35). The DBD of human FOXM1 (amino acids 222–360) was cloned to pET-28a vector, expressed and purified as FOXC2-DBD.

Duplex DNA preparation

All single-stranded DNA oligonucleotides were purchased from Genewiz (SuZhou, China). Complementary oligonucleotides were annealed and purified as described previously (36). The DBE2 contains strands 5′-CAAAATGT AAACAAGA-3′ and 5′-TCTTGTTTACATTTTG-3′ (30). The PC DNA of the palindromic sequence was 5′-ACAC AAATATTTGTGT-3′. DNA3 sequence was 5′-ATTTGT GTACACAAAT-3′. For ITC measurements, all DNA motifs had the same flanking sequence with DBE2 (GTAAAC A site). The other two motifs were as follows: GTACACA site: (5′-TGCAAAATGTACACAAGACT-3′), ACAAAT A site: (5′-TGCAAAATACAAATAAGACT-3′).

Crystallization

FOXC2-DNA complexes were prepared by mixing FOXC2-DBD protein with DBE2 at 1:1.2 molar ratio or with PC at 5:3 molar ratio. The final concentration for crystallization was 10 mg/ml. Crystals of FOXC2-DBD/DBE2 were grown by hanging drop vapor diffusion in 3 days at 4°C using a well solution containing 50 mM Bis–Tris propane (pH 6.68), 14% PEG4K (w/v), 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP. Crystals of FOXC2-DBD/PC DNA were grown under 50 mM sodium acetate (pH 4.7), 14% PEG4K (w/v), 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP. For data collection, single crystal was soaked in well solution plus 20% glycerol (v/v) and flash-frozen in liquid nitrogen for cryo-crystallography. FOXC2-DBD/DBE2 crystal data were collected in our lab. FOXC2-DBD/PC crystal data were collected in BL17U1beamline of Synchrotron Radiation Facility (SSRF) at Shanghai.

Data collection and structure determination

The structures were solved by molecular replacement (MR) using phaser from PHENIX package (37). The recently solved FOXA2/DBE2 structure (PDB ID: 5x07) was used as a search model to solve the FOXC2/DBE2 and FOXC2/PC structures, which takes the space group $P2_12_12_1$, by molecular replacement using Phaser (the sequence identity between the forkhead domains of FOXA2 and FOXC2 is 60%). The structures were then refined in PHENIX. Models were initially refined with simulated annealing, XYZ positional refinement, group B-factor. For late stages of refinement, individual atomic *B*-factors were refined, and solvent or ion molecules were added to the model. The statistics and geometries of the two refined structures are presented in Table 1. Figures were generated using Pymol (38).

DNA structure analysis

DNA structure of unbound DBE2 was predicted by DNAshape (39). Structures of DNA bound to FOXC2, FOXO1 (PDB: 3CO7) and FOXA2 (PDB: 5x07) were analyzed with CURVES (40). For comparison, we renamed the FOXO1 and FOXA2 chains to align the structures and perform binding site analysis in the same strand orientation.

Electrophoretic mobility shift assay (EMSA)

Binding reactions were performed in a total volume of 6 μl in 20 mM HEPES (pH 7.5), 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP. Both protein and DNA were prepared at the concentration of 45 μM. DNA was incubated with protein at room temperature for 20 min. To resolve the free DNA from the protein/DNA complex, we used a native 8% (w/v) polyacrylamide gel in 0.5× TBE buffer. The gel was visualized using 0.5 μg/ml goldview.

Table 1. Data collection and refinement statistics for FOXC2/DNA complexes

	FOXC2/DBE2 DNA	FOXC2/PC16 DNA
Data collection		
Wavelength (Å)	1.5870	0.97915
Space group	*P* 21 21 21	*P* 21 21 21
a, *b*, *c* (Å)	42.54, 72.92, 99.05	41.84, 80.21, 101.15
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00–2.40 (2.48–2.40)	50.00–2.32 (2.41–2.32)
Unique reflections	11 726 (780)	15 111 (1366)
Rsym or Rmerge	0.04 (0.30)	0.13 (0.92)
Redundancy	12.10 (9.70)	13.40 (10.70)
Completeness (%)	92 (63)	98 (92)
Mean I/sigma(I)	29.51 (4.97)	11.30 (2.80)
Refinement		
Resolution (Å)	36.46–2.40	32.24–2.32
Wilson *B*-factor	52.11	57.45
R-work / *R*-free	0.23/0.25	0.22/0.25
Number of non-hydrogen	1473	1464
Macromolecules	1462	1447
Protein residues	97	94
RMS(bonds)	0.006	0.005
RMS(angles)	0.69	0.73
Ramachandran favored (%)	95	99
Rotamer outliers (%)	1.1	2.3
Average *B*-factor	70.16	82.90
Macromolecules	70.27	83.10
Ligands	61.03	82.61
Solvent	55.86	64.15

Isothermal titration calorimetry assays (ITC)

To measure the binding affinities of DNAs with FOX-DBD, ITC measurements were performed (41). The purified protein sample was dissolved in ITC buffer (20 mM HEPES (pH 7.0), 300 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP) at 20–50 μM, and the duplex DNAs were dissolved in ITC buffer at 80∼200 μM. To remove any protein precipitate, the samples were centrifuged before the experiments. All measurements were performed at 25°C using a NANO ITC (TA Instruments). Duplex DNAs were titrated into the experimental cell that contained FOX-DBD protein in twenty-five 2 μl-injection increments. The titration data were analyzed using the launch NanoAnalyze software.

Bioinformatics analysis

Briefly, reads from the FOXC1 IIBA ChIP-seq data (ERP006190) were mapped to mouse genome (GRCm38/mm10) using BFAST (42). Then peaks were called by using MACS2 (43) and the peak sequences were extracted. The core sequences RYAAACA and RYACACA were searched in the ChIP-seq peak sequences.

Site-directed mutagenesis

All mutants of FOXC2-DBD were generated according to the QuikChange mutagenesis protocol (Stratagene, La Jolla, CA, USA) using the pET-28a-FOXC2 plasmid as the template. These mutants were confirmed by DNA sequencing (Genewiz, Suzhou, China). These mutant proteins were expressed and purified as wild type.

Differential scanning fluorimetry assay

To detect the stability of FOXC2-DBD wild type and mutants, a differential scanning fluorimetry (DSF) assay was carried out using a LightCycler 480 real-time PCR instrument (Roche, Switzerland) (44). SYPRO Orange (Sigma, USA) was used as the fluorescent dye, which contacts with the hydrophobic portion of a protein to greatly enhance its fluorescence emission. The proteins (final concentration of 1.0 mg/ml) and dye (final concentration of 5.0 mg/mL) were mixed, and then was sealed and placed in the instrument. To measure melting curves of samples, the temperature was increased gradually with a heating rate of 1.8°C/min. The fluorescence intensity from each tube versus temperature (melt curve) was measured, and a melting temperature (T_m) was calculated from the maximum value of the first derivative of the melt curve. The data were analyzed using the LightCycler 480 software and graphics were produced using the program Excel.

RESULTS

Overall structures of the FOXC2-DNA complex

Most FOX proteins have been reported to bind the canonical DNA response element 5′-RYAAAYA-3′ (R = A/G, Y = C/T) (45,46). However, the molecular mechanism by which FOXC protein binds DNA is not well understood. To address this question, we determined two crystal structures of FOXC2-DBD (residues 72–171) bound to two different 16-bp double-stranded DNAs at 2.40 and 2.32 Å respectively (Table 1 and Figure 1). One contained the Daf-16 family binding element 2 (DBE2) with the recognized consensus motif GTAAACA. The other DNA is derived

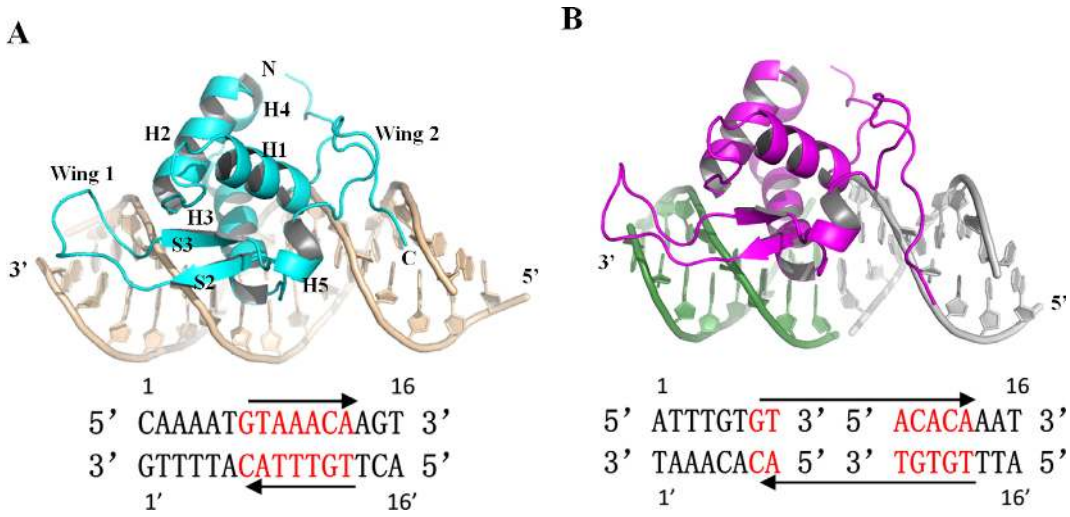


Figure 1. Overall structures of FOXC2-DBD/DNA complex. (**A**) Overall structure of FOXC2-DBD/DBE2 complex. FOXC2-DBD is colored in cyan, and DBE2 is colored in wheat. Secondary structure elements are labeled. The sequence of the 16-bp DNA duplex DBE2 used in crystallization is listed below. (**B**) Overall structure of FOXC2-DBD/DNA3 complex. FOXC2-DBD is colored magenta, and DNA targets are colored in forest green and gray in the symmetry unit. The pseudocontinuous 16-bp DNA sequence is listed below, and the core sequences are colored red.

from the promoter of protein C gene with a putative forkhead motif ACAAATA, which was reported as a recognition site for FOXA1 (47). Both DNAs contain the canonical response element 5′-RYAAAYA-3′ (R = A/G, Y = C/T).

The FOXC2-DBD in both structures displayed a winged-helix fold, composed of three stacking helices (H1, H2, H3), three strands (S1, S2, S3), two wings and two additional small 3_{10}-helices (H4, H5) (Figure 1 and Supplementary Figure S1). The H4 helix connected H2 and H3 as seen in several other FOX proteins (31,32,48–50). The H5 helix within the C-terminal region was followed by a coiled wing 2 (Figure 1). Previous reported structures of FOXP2 (49), FOXD3 (50) and FOXK1a (51) have also shown a C-terminal helix conformation, but the length and sequences are quite different. The superposition of the two FOXC2-DBD structures gave a 0.18 Å root mean square deviation (RMSD) for Cα atoms, indicating little structural difference in these two protein/DNA complexes (Supplementary Figure S2). When our structure of FOXC2-DBD was superimposed with the previous NMR structure in the absence of DNA (33), the RMSD of Cα atoms was ∼0.98 Å. The structural variations mainly locate at wing 1 and the C-terminal region, where conformations were highly disordered without DNA in the NMR structure (Supplementary Figure S2).

In the structure of FOXC2/DBE2 complex (Figure 1A), the recognition helix H3 docked into the major groove roughly perpendicular to the DNA axis and made numerous interactions with core sequence GTAAACA. The typical wing 1, between the S2 and S3, interacted with the phosphate group of the 3′ flanking region of the consensus sequence. The wing 2 interacted with the minor groove of DNA. Surprisingly, the FOXC2-DBD did not bind the ACAAATA sequence in the other complex structure. Instead, FOXC2 bound sequence GTACACA at the pseudocontinuous DNA helix in the symmetry unit by a slightly different interaction pattern (for comparable description, the 16-bp pseudocontinuous DNA sequence ATTTGTGTACACAAAT was named DNA3, Figure 1B).

DNA recognition in FOXC2/DBE2 structure

DNA recognition by FOXC2 was dominantly mediated by the helix H3 as observed in other FOX proteins. Residues Asn118, Arg121, His122 and Ser125 within helix H3 made extensive contacts with the major groove through hydrogen bonds and van der Waals contacts (Figure 2A, Supplementary Figure S3A). Asn118 recognized A10 through bidentate hydrogen bonds. The side chain of His122 protruded into the major groove and formed hydrogen bonds with the bases of T8 and T9′. The main chain and side chain atoms of Arg121 made extensive van der Waals contacts with the bases of T11′, G12′ and T13′. Besides base recognition, Ser125 bound the phosphate backbone of T11′ to further stabilize the complex structure.

In addition to major groove recognition, wing 1 of FOXC2-DBD contributed to DNA binding through interacting with the phosphate backbone at the minor groove of the 3′ flanking core sequence (Figure 2B, Supplementary Figure S3A). Both Ser144 and Trp146 recognized the phosphate group of G12′ through hydrogen bonds. At the stem of wing 1, Lys132 contacted the phosphate groups of T11′. However, the electron density for residues 136–141 and the side chain of Lys142 within wing 1 was not clear in the FOXC2-DBD/DBE2 structure. Compared to helix H3, wing 1 of FOXC2-DBD lacked direct roles in specific base recognition; it helped to stabilize DNA binding.

The C-terminal region of FOXC2-DBD in complex structure consists of a charged wing 2 and a helix H5. Tyr153 within the H5 formed an intra-molecular hydrogen bond with Gln86 of H1 to stabilize the C-terminal conformation (Figure 2C). The basic residues (RRRR) at the end of wing 2 pointed toward the DNA minor groove and contacted the phosphate backbones of upstream nucleotides from the core sequence. Specifically, the guanidino group

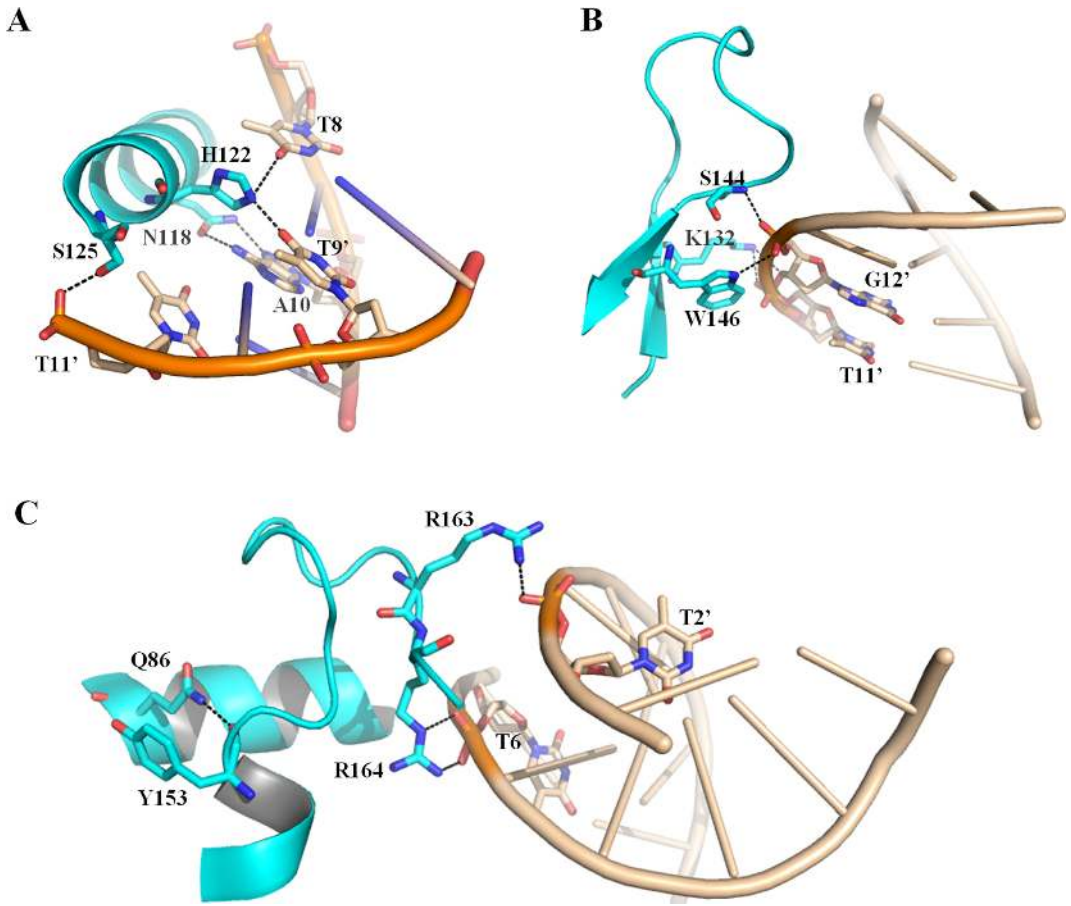


Figure 2. Detailed contacts in the FOXC2-DBD/DBE2 interface. (**A**) Recognition of the GTAAACA core sequence by helix H3 of FOXC2-DBD. (B, C) Hydrogen bond interactions between DNA and wing 1 (**B**) or C terminus (**C**) of FOXC2-DBD. All hydrogen bonds are shown as black dotted lines.

of Arg163 and Arg164 formed hydrogen bond and van der Waals contacts with phosphate groups of T2′ and T6, respectively (Figure 2C). Similar to wing 1, wing 2 of FOXC2-DBD was not involved in base-specific interaction, but stabilized the DNA binding through its interaction with DNA phosphate backbones.

The FOXC2-DBD/DBE2 structure showed that residues from the N-terminus were also involved in DNA binding. Residues Lys72 and Ser76 respectively formed direct hydrogen bonds or van der Waals contacts with the phosphate backbones (Supplementary Figure S3). In addition, these contacts were further stabilized by the hydrogen bond between residue Tyr77 from helix H1 and the phosphate group of T8 and G7. Therefore, helix H3 of FOXC2 provides all of the base-specific contacts with DBE2, while the N-terminus, wing 1 and C-terminus of FOXC2-DBD make additional contacts with the phosphate groups.

FOXC2 bound DNA containing GTACACA motif

In the other crystal structure of FOXC2-DBD/DNA3 complex, FOXC2 bound the core sequence GTACACA at the pseudocontinuous DNA helix in the symmetry unit. The contact pattern of the major groove recognition is different from that of DBE2. Similarly, H3 docked into the major groove of the pseudocontinuous DNA and contributed the major base contacts. Interestingly, the residue Asn118, which formed bidentate hydrogen bonds with A10 in the FOXC2/DBE2 structure (Figure 3B), did not form any hydrogen bond with DNA in the FOXC2/DNA3 structure (Figure 3A). On the other hand, the residue Arg121, which did not form any hydrogen bond with DNA in the FOXC2/DBE2 structure, made a hydrogen bond with the base of G12′ in the FOXC2/DNA3 structure (Figure 3). The residue His122 retained its interaction with T8 through a hydrogen bond in both structures (Supplementary Figure S3). Apart from the differences in direct base recognition, the backbone interactions also had some deviations between the two structures (Supplementary Figure S3).

Arg121 seemed to play key roles in the maintenance of local protein conformation, as well as DNA recognition by FOXC2 protein. In both structures, the side chain of Arg121 was stabilized by the side chains of Tyr99 and Asn118 through direct hydrogen bonds (Figures 3C and D). As for DNA recognition, Arg121 made a hydrogen bond with the base of G12′ and extensive van der Waals contacts with the bases of T11′, G11′, and T13′ in the FOXC2/DNA3 structure (Figure 3C); in the FOXC2/DBE2 structure, although Arg121 did not form any hydrogen bond with the base of G12′, it made extensive van der Waals contacts with the bases of G12′ and its neighboring T11′ and T13′ (Figure 3D).

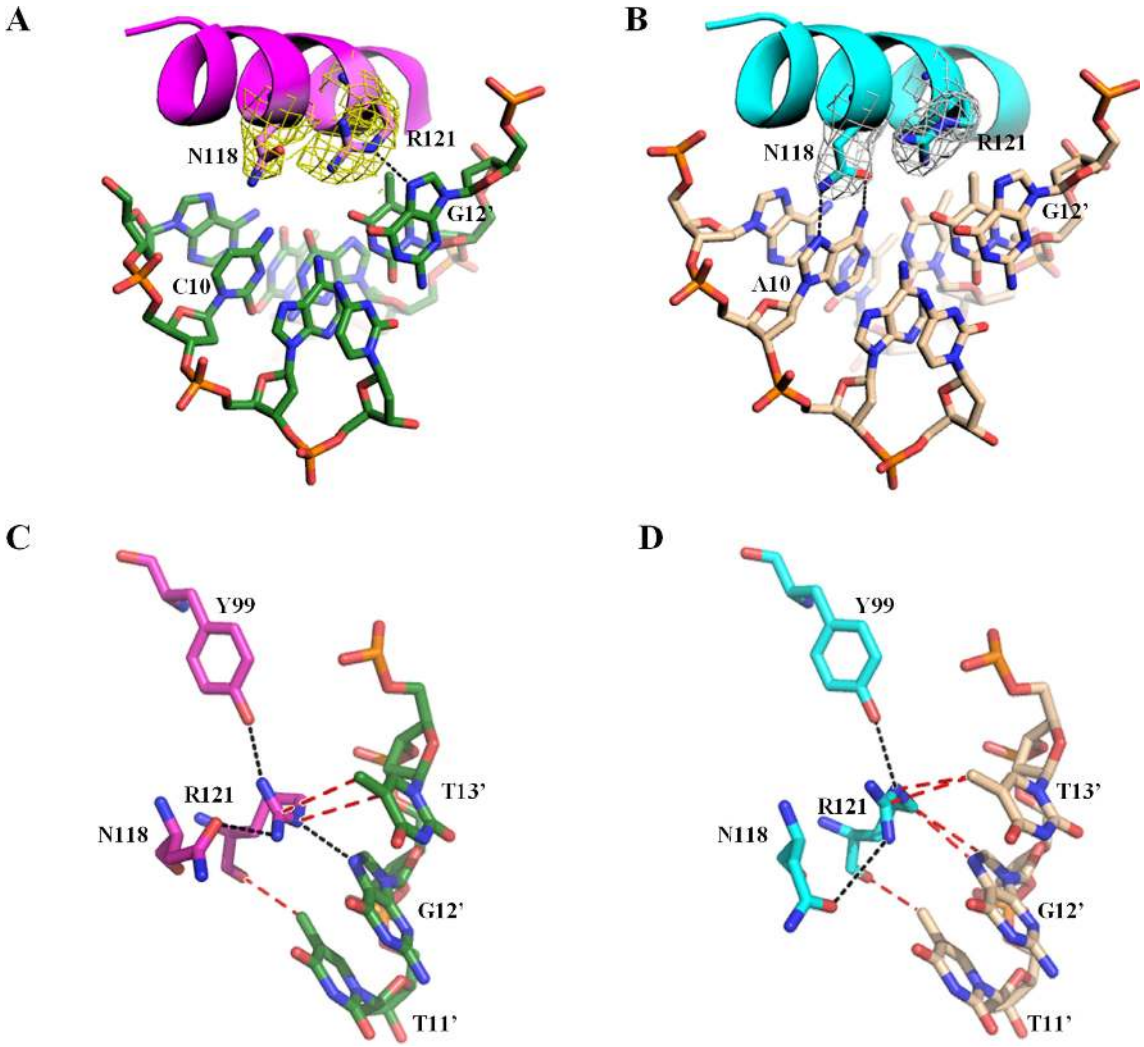


Figure 3. The comparison of DNA recognition by FOXC2-DBD in different structures. (**A**) R121 forms a hydrogen bond with the base of G12′ in the FOXC2/DNA3 structure. (**B**) N118 forms bidentate hydrogen bonds with the base of A10 in the FOXC2/DBE2 structure. (**C**) Role of R121 in the FOXC2/DNA3 structure. (**D**) Role of R121 in the FOXC2/DNA3 structure. The $2F_o - F_c$ electron density map contoured at 1.0σ is colored yellow in the FOXC2/DNA3 structure, and gray in the FOXC2/DBE2 structure. FOXC2-DBD and DNAs are colored as in Figure 1. Hydrogen bonds are shown as black dotted lines, and van der Waals contacts are shown as red dashed lines.

DNA binding specificities of forkhead domains

To facilitate our understanding of DNA binding specificities by the highly conserved forkhead recognition helix, we systematically analyzed structures of FOXC2 and other FOX proteins bound to various sequences. The major contributors for FOX proteins to make direct base-specific interactions with target DNA are three nearby conserved residues in H3. We termed it 'N(S/A)IRH' motif here (Supplementary Figure S1). We analyzed the DNA recognition pattern of this motif in FOXC2, and many other FOX proteins, such as FOXM1 (32), FOXK1a (51) and so on (29–31,51,52). In all structures, the histidine within this motif forms hydrogen bonds with DNA bases in a similar manner. However, the contribution of the asparagine and arginine residues could be classified into three categories. (i) In FOXC2/DBE2 structure, the asparagine forms bidentate hydrogen bonds with TAAACA motif (the interacting nucleotide is underlined) (Figure 3A). The same bidentate hydrogen bonds are also formed in the case of FOXO1 bound to TAAATC (5), and FOXP3 bound to CAAATT (53) (Figure 4A). (ii) If this adenine is substituted by other nucleotides, this bidentate hydrogen bond will not form, as in the case of FOXC2 bound to TACACA (Figure 4B, left) and FOXK1a bound to AATACA (Figure 4B, right) (51). Under this circumstance, the arginine within the 'N(S/A)IRH' motif will make hydrogen bond contact with G12′ of the complementary chain (Figure 4B, left). Similarly, the corresponding arginine of FOXK1a-DBD2 plays the same role in AATACA recognition (Figure 4B, right). (iii) In some cases, both asparagine and arginine within the 'N(S/A)IRH' motif contribute to hydrogen bonding with DNA (Figure 4C). In conclusion, although most FOX proteins have been reported to bind RYAAAYA motif preferentially, FOX proteins may bind similar sequences with flexibility.

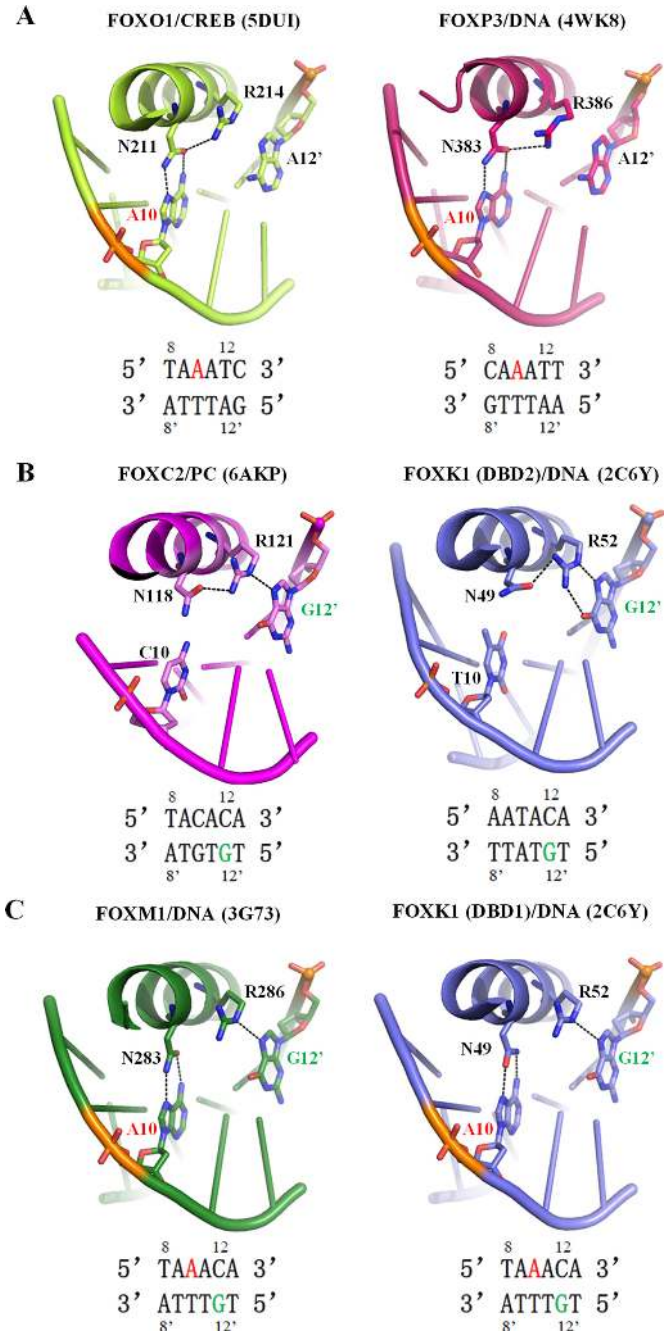


Figure 4. DNA base recognition by FOX proteins. Asparagine and Arginine of the 'N(S/A)IRH' motif involved in various DNA core sequence recognition are shown as stick and labeled. The DNA binding motif is listed below, with A10 colored red and G12′ colored green. Hydrogen bonds are shown as black dotted lines.

Comparison with FOXA2/DBE2 and FOXO1/DBE2 structures

Above we described the structural basis of FOXC2-DBD binding to two different DNA sequences, and analyzed the specificity of the recognition helix bound to varying core sequences. Then we tried to compare the structural variations of different FOX proteins bound to the same DNA sequence. We superimposed the FOXC2/DBE2 structure with the previously reported FOXO1/DBE2 and FOXA2/DBE2 structures. They share similar overall winged-helix structures with RMSDs for Cα atoms of about 0.44 and 0.52 Å, respectively (Figure 5A). The three FOX proteins show similar manners of specific base interactions. The conserved asparagine and histidine form direct hydrogen bonds with the core sequence. Some other residues devote numerous van der Waals contacts to the target DNA. A major difference in DNA binding lies in backbone interactions with the 5′ flanking core sequence. N-terminal residue Lys72 of FOXC2 makes a direct hydrogen bond with G7 (Figure 5C). FOXO1 also utilizes Arg157 and Asn158 in the N-terminal tail to interact directly with G7 (Figure 5D). Lys159 of FOXA2, which shares the same sequence with FOXC2 (Supplementary Figure S1), is not involved in direct DNA binding. Nevertheless, helix H3 residue Ser206 of FOXA2 forms additional hydrogen bond contact with phosphate groups of T8 (Figure 5E). Another major structural deviation maps to the C-terminal region. The wing 2 region of FOXA2 and FOXO1 are flexible and not observed in previously solved structures (Figure 5A), though the truncated mutation assay showed that the wing 2 was essential for DNA binding (30,35). However, the C terminus of FOXC2-DBD has well defined electron density that allows for clear construction of the model in this area (Supplementary Figure S4). The structure exhibits that the guanidino group of Arg163 and Arg164 make direct hydrogen bonds with the phosphate backbones of T2′ and T6, respectively (Figure 2C, Figure 5C). Wing 1 region of these three proteins exhibits slightly different contacts with the backbone of the 3′ flanking region, corresponding to a small DNA deviation.

DNA structure and deformability may also contribute to binding specificity (54), so we also analyzed the DNA structural features for the three complexes, as well as unbound DBE2 B-DNA. The three FOX proteins induce somewhat different deformation on the same DNA sequence (Figure 5A, B). Compared with standard B-DNA conformation, the minor groove width of DBE2 bound by FOX proteins widens at the center of the core binding site (Figure 5B). DBE2 bound by FOXC2 and FOXO1 has its maximal value of minor groove width at T8, while DBE2 bound by FOXA2 has its maximal minor groove width shifted by one nucleotide position (at A9). DNA deformation induced by each of the FOX proteins deviates most from each other at the 5′ flanking sequence (T6 to T8). This might be related to the different backbone interactions at this region (Figure 5C, D). In addition, this deviation might, in some cases, be due to crystal packing.

DNA binding affinity

To further characterize the DNA-binding properties of FOXC2-DBD, we tested its binding ability to several different DNA sequences (Figure 6). Three sequences were tested: DBE2 containing GTAAACA core sequence, PC containing the element ACAAATA, and DNA3 containing motif GTACACA. First, the capacity to form DNA-protein complexes were determined using electrophoretic mobility shift assays (EMSAs). As shown in Figure 6A, FOXC2-DBD bound to both DBE2 and DNA3, with slightly better

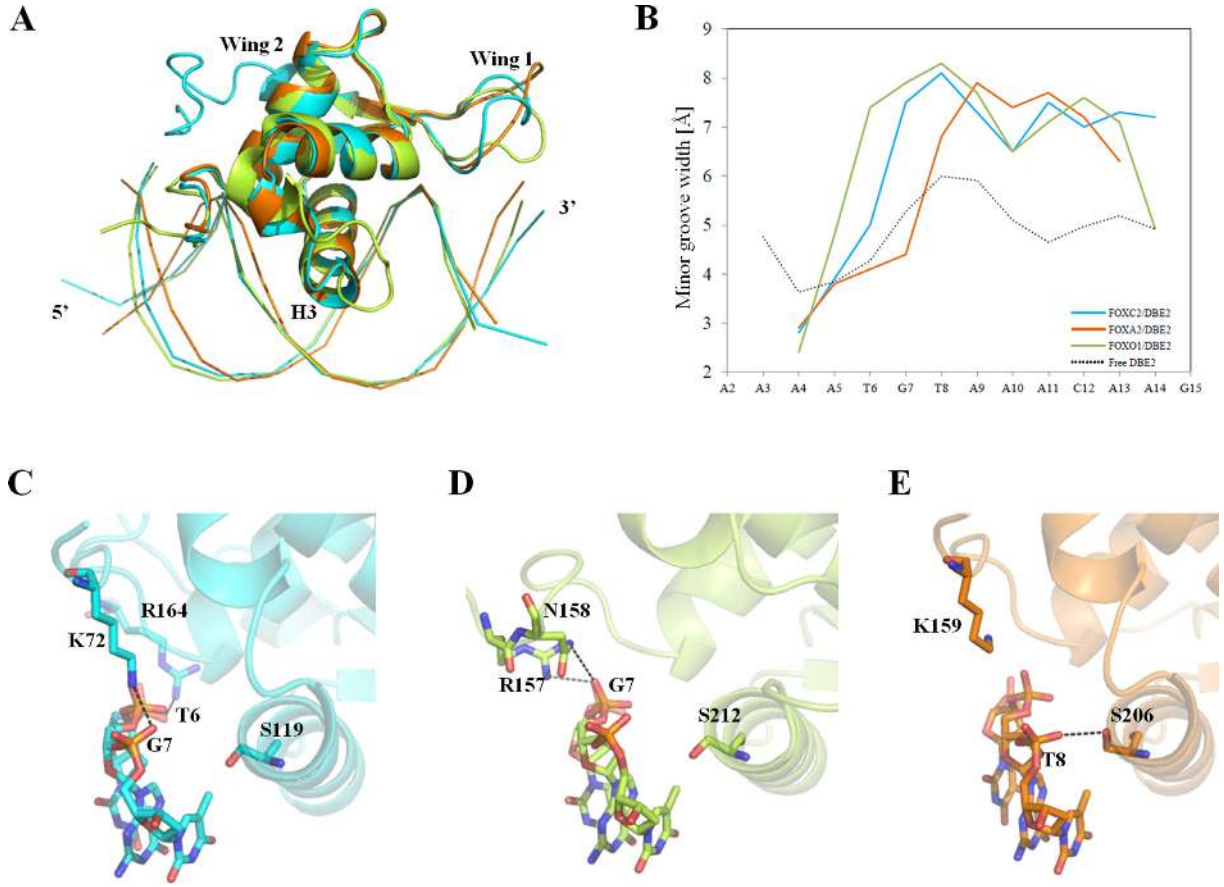


Figure 5. Structural comparison of FOXC2-DBD/DBE2 with FOXO1-DBD/DBE2 (PDB: 3CO7) and FOXA2-DBD/DBE2 (PDB: 5x07). (**A**) Superposition of FOXC2 structure (cyan) with previous FOXO1 (green) and FOXA2 (orange). (**B**) DNA minor groove widths of bound and unbound DNAs. DBE2 bound by FOXC2 is shown as cyan line. DBE2 bound by FOXA2 is shown as green line. DBE2 bound by FOXA2 is shown as orange line. Unbound DBE2 is shown as a dotted brown line. (C–E) Detailed comparison of the interactions between FOXC2 (**C**), FOXO1 (**D**) and FOXA2 (**E**) with DNA.

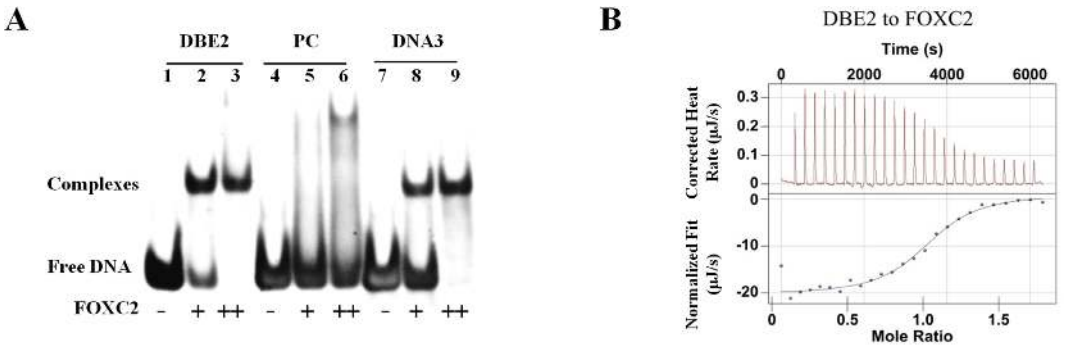
affinity with DBE2. However, FOXC2-DBD showed only very weak binding with sequence PC, with a higher shift. The higher shift could be due to the presence of a palindromic sequence 'ACAAA*TATTTGT*' with two putative forkhead binding sites (one site underlined, the other site in italics).

To quantitatively analyze DNA binding affinity with these three DNA binding motifs (DBE2, DNA3 and DNA4), isothermal titration calorimetry (ITC) was carried out. Since sequence PC have two putative forkhead binding sites, we replaced sequence PC with DNA4 (containing only one 'ACAAATA' site). The representative binding isotherm of DBE2 (containing GTAAACA motif) is presented in Figure 6B. The K_d values of FOXC2-DBD bound to GTAAACA and GTACACA were estimated to be 0.79 and 2.22 μM, respectively (Figure 6C). An estimated K_d value of FOXC2-DBD toward ACAAATA was larger than 100 μM. These results indicated that FOXC2-DBD bound GTACACA with a 3-fold lower affinity than that of GTAAACA, while it could hardly bind to the ACAAATA site.

In order to provide more mechanistic insights into the basis for sequence specificity of different FOX proteins, we studied the DNA binding affinity of two other FOX proteins (FOXA2 and FOXM1) towards motifs GTAAACA, GTACACA and ACAAATA. The results are shown in Figure 6C. Similar to FOXC2-DBD, proteins FOXA2-DBD and FOXM1-DBD bound motif GTAAACA with high affinity (0.29 and 2.45 μM, respectively), and bound motif GTACACA with 3–4 folds lower affinity than motif GTAAACA (0.99 and 8.3 μM, respectively). However, For ACAAATA site, these FOX proteins exhibit quite different binding affinity. FOXA2 showed a 26-fold lower affinity (7.54 μM) than motif GTAAACA (0.29 μM), while FOXC2 could barely bind motif ACAAATA. On the other hand, FOXM1 had a similar affinity with the ACAAATA site (2.07 μM) as the GTAAACA site (2.45 μM). These results suggest that the RYAAAYA canonical motif does not apply to all FOX proteins.

Identification of endogenous FOXC binding motif RYAAACA and RYACACA

Compared to the canonical forkhead binding motif RYAAAYA (R = G/A, Y = T/C), FOXC2 seems to prefer a C at the sixth position, while it is flexible at the fourth position, in our biochemical studies. In order to determine whether FOXC protein has the same binding preference *in vivo*, we tried to look for the occurrence of motifs RYAAACA or RYACACA in the ChIP-seq database. How-



C

DNA	Binding site	ITC parameters	FOXC2	FOXA2	FOXM1
DBE2	GTAAACA	Kd (μM)	0.79±0.18	0.29±0.16	2.45±0.63
		n	1.03	1.04	0.99
		△H (KJ/mol)	-21.60±2.18	-27.31±1.23	-7.22±0.31
		△S (J/mol·K)	44.43±2.64	32.05±1.67	83.50±2.00
DNA3	GTACACA	Kd (μM)	2.22±0.32	0.99±0.23	8.30±1.49
		n	1.06	1.11	1.05
		△H (KJ/mol)	-22.48±0.45	-26.05±1.25	-14.61±0.51
		△S (J/mol·K)	32.84±3.66	26.92±1.46	48.39±3.20
DNA4	ACAAATA	Kd (μM)	>121.4±4.56	7.54±0.12	2.07±0.46
		n	0.94	1.09	1.02
		△H (KJ/mol)	97.54±5.23	16.21±1.19	-5.52±0.16
		△S (J/mol·K)	402.1±5.88	43.72±2.45	90.54±1.99

Figure 6. DNA binding affinities of FOX proteins. (**A**) DNA-binding features of FOXC2-DBD are measured by EMSA. (**B**) Measurement of DNA binding affinity by quantitative ITC. The graph represents the curve of FOXC2-DBD titrated by DBE2. Representative power-response curves (top) and heats of reaction normalized to the moles of protein injected (bottom). (**C**) Thermodynamic parameters of the FOX/DNA interactions. Values represent mean and standard deviation of at least three independent experiments.

ever, there is no FOXC2 ChIP-seq data available. Given the high identity between FOXC2 and FOXC1, we analyzed a deposited FOXC1 ChIP-seq dataset (ERP006190). For motif RYAAACA, 7721 peaks were found (Figure 7A), accounting for 32.3% of total FOXC1 binding sites. It was identified in FOXC1 binding sites near the promoters of genes such as *Cd36*, *Sirt1*, *Per1*, *Akt1*, *Igf1* (Figure 7B). For motif RYACACA, 3999 peaks were found, accounting for 16.7% of total FOXC1 binding sites (Figure 7A). Specifically, motif RYACACA was identified in FOXC1 binding sites near the promoters of genes such as *Mdm4*, *Atf6*, *Fgf7*, *Yes1*, *Foxp2* (Figure 7B). These data suggested that FOXC protein indeed could bind both RYAAACA and RYACACA sites *in vivo*.

Disease-causing mutations

FOXC1 and FOXC2 mutations are associated with ARS or LDS, and most mutations locate in the DNA-binding domain (Figure 8A). Since the DNA-binding domain of FOXC2 and FOXC1 share 98% sequence identity (Supplementary Figure S1), we also mapped the FOXC1 mutations on FOXC2 structure. These mutations include R121H and S125L in helix H3, C129Y, K132E, G143D and W146G in wing 1 region, R163P and R164W in C-terminal region (corresponding to R127H, S131L, C135Y, K138E, G149D, R169P and R170W in FOXC1) (Figure 8B).

We first performed differential scanning fluorimetry (DSF) to analyze the protein stability. The melting Curves of wild type FOXC2-DBD and its mutants are shown in Figure 8C. The melting temperature (T_m) of wild type FOXC2 was 56.1°C. Three mutants, R121H, C129Y, and R163P showed lower T_m than wild type. Among the three mutants, the R121H mutant exhibited the most dramatic decrease of T_m (46.2°C), suggesting that R121H is a highly destabilizing mutation. This result is consistent with our structural observation that Arg121 plays key roles in maintaining local protein conformation.

Then we tested the DNA binding of these mutants using EMSA. As shown in Figure 8D, compared with wild-type FOXC2-DBD, all mutants showed somewhat reduced DNA binding affinities. Three mutants (S125L, K132E and R121H) exhibited greatly reduced DNA binding affinities. This is consistent with our structural observation that

A

Binding sites	total	Sites containing RYAAACA	Sites containing RYACACA	Sites containing both RYAAACA and RYACACA
Number of peaks	23878	7721	3999	1685
Ratio	-	32.3%	16.7%	7.1%

B

Position	RYAAACA		RYACACA	
	Gene	Sequence	Gene	Sequence
Promotor	Cd36	ACTTATGTAAACAAGGTTT	Mdm4	CATTGAGTACACACTAGTC
Promotor	Sirt1	TCAACGACAAACAATAGCA	Atf6	ACCCGCATACACACGTGCA
Promotor	Per1	TTCTGGGTAAACAAGTTGC	Fgf7	CTGTGAGTACACAGTGTTT
Promotor	Akt1	AGGAGTGTAAACAAGAGGG	Yes1	CCATGTGTACACAAGGACA
Promotor	Igf1	TGAAAGGTAAACACTGCTG	Foxp2	AGGCGCACACACAGGAGAG

Figure 7. Occurrence of the FOXC2 binding motif in FOXC1 binding sites in human cells. (**A**) Analysis ratio of motif RYAAACA and RYACACA in FOXC1 ChIP-seq data (ERP006190). (**B**) Representative gene promoters near FOXC1 binding sites containing motif RYAAACA or RYACACA.

Ser125, Lys132 and Arg121 all directly interact with DNA. In particular, the R121H mutant showed no detectable DNA binding. The total loss of DNA binding of the R121H mutant could also be due to the fact that this mutation is a highly destabilizing mutation as shown in our DSF experiments.

DISCUSSION

FOX proteins share conserved DNA binding domains with a winged-helix fold, but have diverse DNA recognition features and functions (1). FOXC subfamily proteins play important roles in the embryonic development and development of multiple organs such as eye, heart, etc (4). Mutations and deregulation of FOXC proteins are associated with developmental abnormalities, and tumors (20,21). In this study, we report the crystal structures of FOXC2-DBD bound to two different DNA sequences, and its binding properties using EMSA and ITC assays.

FOX proteins have conserved amino acid sequences 'N(S/A)IRH' in the recognition helix H3, and recognize a similar DNA sequence binding site (5′-RYAAAYA-3′, R = A/G, Y = C/T) with respective preference nucleotides at the R and Y positions (45,46). FOXC2 recognizes the DBE2 sequence containing consensus sequence GTAAACA at the highest affinity. It also binds sequence GTACACA with a 3-fold lower affinity than that of GTAAACA (Figure 1B, Figure 6C), but can hardly bind to a DNA duplex containing ACAAATA (Figure 6C). It seems that the canonical RYAAAYA (R = G/A, Y = T/C) motif does not fit well with the FOXC2 protein. Based on our results, we define the FOXC2 binding motif as RYAMACA (R = G/A, Y = T/C, M = A/C).

Other FOX proteins, FOXA2 and FOXM1, also showed similar binding patterns to motifs GTAAACA and GTAC ACA (Figure 6C). However, these FOX proteins showed very different binding affinities to ACAAATA site: (i) FOXC2 can hardly bind the ACAAATA site; (ii) FOXA2 can still bind the ACAAATA site, but with a much lower affinity than that of the GTAAACA site; (iii) FOXM1 can bind the ACAAATA site with a similar affinity to that of the GTAAACA site. We carefully examined the protein sequences and crystal structures of these proteins; however, we could not explain the mechanisms by which these forkhead proteins bind very differently to ACAAATA site.

Analysis of FOXC-binding sites using ChIP-seq data revealed that 32.3% of FOXC1-binding sites follow the consensus motif RYAAACA (Figure 7A). The number is ∼2-fold higher than 16.7% for motif RYACACA, consistent with the higher binding affinity of GTAAACA than GT ACACA (Figure 6C). The fact that FOXC can bind motifs GTAAACA and GTACACA with different binding affinities suggests a mechanism of transcriptional regulation by FOXC proteins. At low expression level, FOXC may preferentially bind to the high-affinity site GTAAACA; while at high expression level, FOXC may bind both high-affinity and low-affinity sites. Thus, depending on the expression level, FOXC proteins may regulate different sets of genes, leading to different biological functions.

We noticed that about half of the ChIP sites do not contain the RYAMACA motif (R = G/A, Y = T/C, M = A/C)

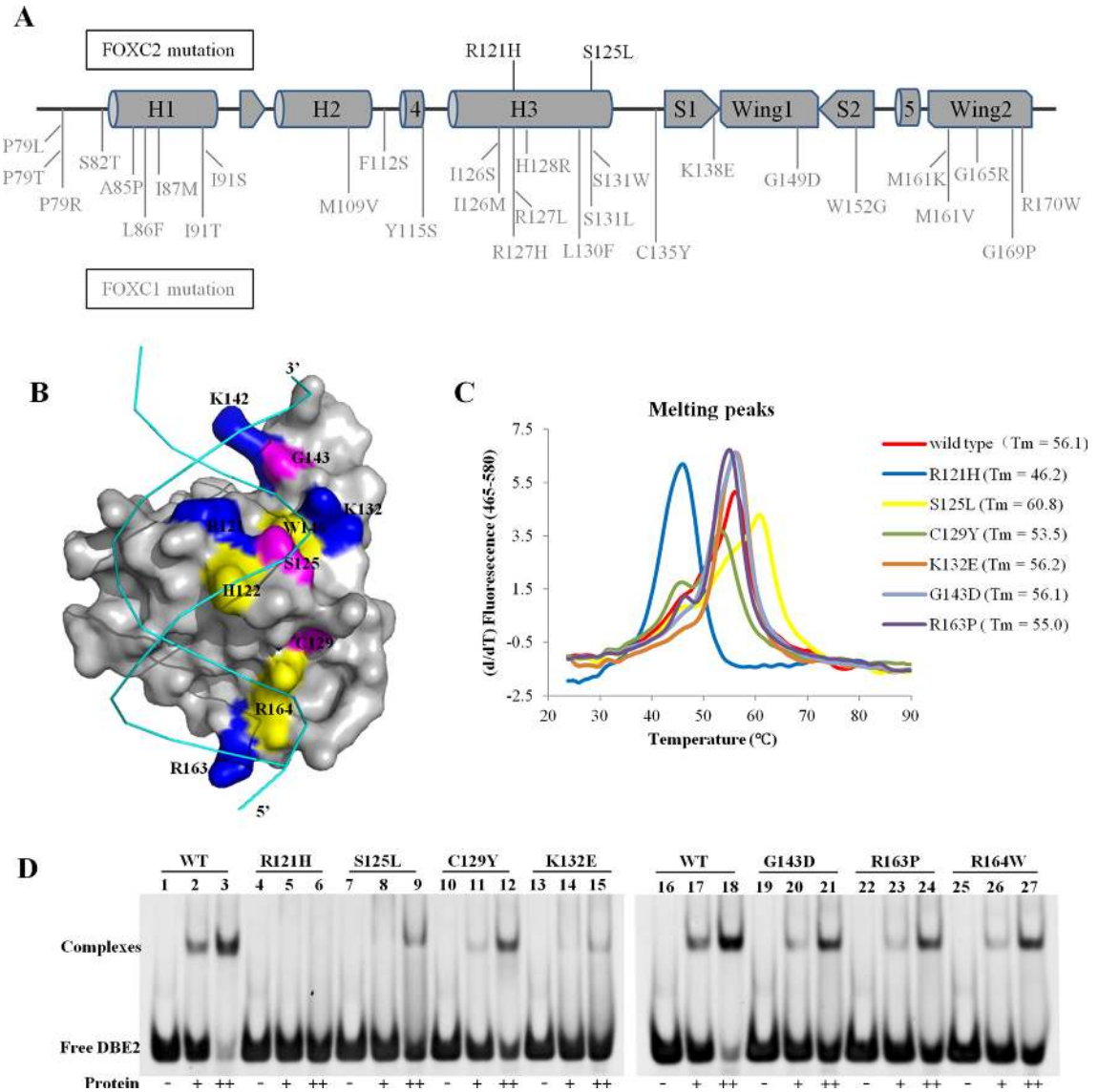


Figure 8. The effect of disease-causing mutations. (**A**) The location of mutations in the DNA binding domain of FOXC1 (gray) and FOXC2 (black). (**B**) Surface representation of FOXC2-DBD bound to DBE2 displaying mutated residues. (**C**) Melting curves of FOXC2 wild-type and mutants in DSF assays. (**D**) DNA binding ability of FOXC2 wild-type and mutants is measured using EMSA.

as we described. There could be several possibilities: (i) in addition to these sequences, FOXC protein could also bind other sequences. For example, FOXO1 and FOXP3 bind GTAAATC and ACAAATT respectively (5,53); FOXA3 binds GTCAACC, and was shown to be similar to globular domain of histone H5 (29). In our study, FOXC2 can bind GTACACA. Moreover, a recent report showed that FOX proteins could bind completely distinct DNA sequences, such as GACGC, GATGC, CGCAC sites (55). From a structural point of view, FOXC proteins use extensive van der Waals contacts and a relatively small number of hydrogen bonds in the major groove, therefore it may bind a variety of sequences as long as the sequences maintain a few hydrogen bond determinants and has shape complementary to the DNA binding surface. (ii) FOXC protein may use domains other than forkhead domain to bind DNA. In addition to the forkhead domain, FOXC protein also has two activation domains and an inhibitory domain. It is possible that these domains can also contribute to binding site selection. iii, FOXC protein could bind DNA indirectly. FOXC1 has been shown to interact with several other transcription factors, such as GLI2 (8), PITX2 (56), and PBX1 (57).

In summary, we have determined crystal structures of two FOXC2-DBD/DNA complexes, and revealed the structural mechanisms by which FOXC2 binds DNA. The helix H3 provides all the base-specific contacts, while the N-terminus, wing 1, and the C-terminus of FOXC2-DBD all make additional contacts with the phosphate groups of DNA. Our structural, biochemical, and bioinformatic analyses also reveal that FOXC2 binds DNA sequences slightly different from the consensus forkhead motif. We define the FOXC binding motif as 'RYAMACA' (R = G/A, Y = T/C, M = A/C). In addition, our crystal structures and accompany-

ing biochemical assays provide a molecular basis for understanding disease-causing mutations in FOXC1 and FOXC2.

DATA AVAILABILITY

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

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff from BL17U beamline of the Shanghai Synchrotron Radiation facility (SSRF) for help with data collection. We thank Dr. Michael R. Stallcup for proofreading.

FUNDING

National Natural Science Foundation of China [81372904 and 81570537 to Y.C., 81272971 to Z.C.]. Funding for open access charge: National Natural Science Foundation of China [81372904 and 81570537 to Y.C., 81272971 to Z.C.].
Conflict of interest statement. None declared.

REFERENCES

1. Golson,M.L. and Kaestner,K.H. (2016) Fox transcription factors: from development to disease. *Development*, **143**, 4558–4570.
2. Katoh,M., Igarashi,M., Fukuda,H., Nakagama,H. and Katoh,M. (2013) Cancer genetics and genomics of human FOX family genes. *Cancer Lett.*, **328**, 198–206.
3. Weigel,D., Jurgens,G., Kuttner,F., Seifert,E. and Jackle,H. (1989) The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the drosophila embryo. *Cell Rep.*, **57**, 645–658.
4. Hannenhalli,S. and Kaestner,K.H. (2009) The evolution of Fox genes and their role in development and disease. *Nat. Rev. Genet.*, **10**, 233–240.
5. Singh,P., Han,E.H., Endrizzi,J.A., O'Brien,R.M. and Chi,Y.I. (2017) Crystal structures reveal a new and novel FoxO1 binding site within the human glucose-6-phosphatase catalytic subunit 1 gene promoter. *J. Struct. Biol.*, **198**, 54–64.
6. Aldinger,K.A., Lehmann,O.J., Hudgins,L., Chizhikov,V.V., Bassuk,A.G., Ades,L.C., Krantz,I.D., Dobyns,W.B. and Millen,K.J. (2009) FOXC1 is required for normal cerebellar development and is a major contributor to chromosome 6p25.3 Dandy-Walker malformation. *Nat. Genet.*, **41**, 1037–1042.
7. Mears,A.J., Jordan,T., Mirzayans,F., Dubois,S., Kume,T., Parlee,M., Ritch,R., Koop,B., Kuo,W.L., Collins,C. *et al.* (1998) Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly. *Am. J. Hum. Genet.*, **63**, 1316–1328.
8. Yoshida,M., Hata,K., Takashima,R., Ono,K., Nakamura,E., Takahata,Y., Murakami,T., Iseki,S., Takano-Yamamoto,T., Nishimura,R. *et al.* (2015) The transcription factor Foxc1 is necessary for Ihh-Gli2-regulated endochondral ossification. *Nat. Commun.*, **6**, 6653.
9. Kume,T., Deng,K.Y., Winfrey,V., Gould,D.B., Walter,M.A. and Hogan,B.L. (1998) The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. *Cell*, **93**, 985–996.
10. Cederberg,A., Gronning,L.M., Ahren,B., Tasken,K., Carlsson,P. and Enerback,S. (2001) FOXC2 is a winged helix gene that counteracts obesity, hypertriglyceridemia, and diet-induced insulin resistance. *Cell*, **106**, 563–573.
11. You,W., Fan,L., Duan,D., Tian,L., Dang,X., Wang,C. and Wang,K. (2014) Foxc2 over-expression in bone marrow mesenchymal stem cells stimulates osteogenic differentiation and inhibits adipogenic differentiation. *Mol. Cell. Biochem.*, **386**, 125–134.
12. Wu,X. and Liu,N.F. (2011) FOXC2 transcription factor: a novel regulator of lymphangiogenesis. *Lymphology*, **44**, 35–41.
13. Tsuji,M., Morishima,M., Shimizu,K., Kume,T., Tamaoki,J. and Ezaki,T. (2014) The role of Foxc2 gene in lung development. *Eur. Respir. J.*, **59**, 501–514.
14. Kume,T., Jiang,H., Topczewska,J.M. and Hogan,B.L. (2001) The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.*, **15**, 2470–2482.
15. Seo,S. and Kume,T. (2006) Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. *Dev. Biol.*, **296**, 421–436.
16. Mayeuf-Louchart,A., Montarras,D., Bodin,C., Kume,T., Vincent,S.D. and Buckingham,M. (2016) Endothelial cell specification in the somite is compromised in Pax3-positive progenitors of Foxc1/2 conditional mutants, with loss of forelimb myogenesis. *Development*, **143**, 872–879.
17. Omatsu,Y., Seike,M., Sugiyama,T., Kume,T. and Nagasawa,T. (2014) Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation. *Nature*, **508**, 536–540.
18. Wang,L., Siegenthaler,J.A., Dowell,R.D. and Yi,R. (2016) Foxc1 reinforces quiescence in self-renewing hair follicle stem cells. *Science*, **351**, 613–617.
19. Wei,C., Lin,H. and Cui,S. (2018) The forkhead transcription factor FOXC2 is required for maintaining murine spermatogonial stem cells. *Stem Cells Dev.*, **27**, 624–636.
20. Wang,T., Zheng,L., Wang,Q. and Hu,Y.W. (2018) Emerging roles and mechanisms of FOXC2 in cancer. *Clin. Chim. Acta*, **479**, 84–93.
21. Han,B., Bhowmick,N., Qu,Y., Chung,S., Giuliano,A.E. and Cui,X. (2017) FOXC1: an emerging marker and therapeutic target for cancer. *Oncogene*, **36**, 3957–3963.
22. Elian,F.A., Yan,E. and Walter,M.A. (2018) FOXC1, the new player in the cancer sandbox. *Oncotarget*, **9**, 8165–8178.
23. Hollier,B.G., Tinnirello,A.A., Werden,S.J., Evans,K.W., Taube,J.H., Sarkar,T.R., Sphyris,N., Shariati,M., Kumar,S.V., Battula,V.L. *et al.* (2013) FOXC2 expression links epithelial-mesenchymal transition and stem cell properties in breast cancer. *Cancer Res.*, **73**, 1981–1992.
24. Kume,T. (2008) Foxc2 transcription factor: a newly described regulator of angiogenesis. *Trends Cardiovasc. Med.*, **18**, 224–228.
25. Yu,M., Bardia,A., Wittner,B.S., Stott,S.L., Smas,M.E., Ting,D.T., Isakoff,S.J., Ciciliano,J.C., Wells,M.N., Shah,A.M. *et al.* (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*, **339**, 580–584.
26. Xia,L., Huang,W., Tian,D., Zhu,H., Qi,X., Chen,Z., Zhang,Y., Hu,H., Fan,D., Nie,Y. *et al.* (2013) Overexpression of forkhead box C1 promotes tumor metastasis and indicates poor prognosis in hepatocellular carcinoma. *Hepatology*, **57**, 610–624.
27. Seifi,M. and Walter,M.A. (2017) Axenfeld-Rieger syndrome. *Clin. Genet.*, **93**, 1123–1130.
28. Bell,R., Brice,G., Child,A., Murday,V., Mansour,S., Sandy,C., Collin,J., Brady,A., Callen,D., Burnand,K. *et al.* (2001) Analysis of lymphoedema-distichiasis families forFOXC2 mutations reveals small insertions and deletions throughout the gene. *Hum. Genet.*, **108**, 546–551.
29. Clark,K.L., Halay,E.D., Lai,E. and Burley,S.K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, **364**, 412–420.
30. Brent,M.M., Anand,R. and Marmorstein,R. (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure*, **16**, 1407–1416.
31. Boura,E., Rezabkova,L., Brynda,J., Obsilova,V. and Obsil,T. (2010) Structure of the human FOXO4-DBD-DNA complex at 1.9 A resolution reveals new details of FOXO binding to the DNA. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 1351–1357.
32. Littler,D.R., Alvarez-Fernandez,M., Stein,A., Hibbert,R.G., Heidebrecht,T., Aloy,P., Medema,R.H. and Perrakis,A. (2010) Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic Acids Res.*, **38**, 4527–4538.
33. van Dongen,M.J., Cederberg,A., Carlsson,P., Enerback,S. and Wikstrom,M. (2000) Solution structure and dynamics of the

DNA-binding domain of the adipocyte-transcription factor FREAC-11. *J. Mol. Biol.*, **296**, 351–359.
34. Wu,D., Guo,M., Philips,M.A., Qu,L., Jiang,L., Li,J., Chen,X., Chen,Z., Chen,L. and Chen,Y. (2016) Crystal structure of the FGFR4/LY2874455 complex reveals insights into the Pan-FGFR selectivity of LY2874455. *PLoS One*, **11**, e0162491.
35. Li,J., Dantas Machado,A.C., Guo,M., Sagendorf,J.M., Zhou,Z., Jiang,L., Chen,X., Wu,D., Qu,L., Chen,Z. *et al.* (2017) Structure of the Forkhead Domain of FOXA2 Bound to a Complete DNA Consensus Site. *Biochemistry*, **56**, 3745–3753.
36. Chen,Y., Bates,D.L., Dey,R., Chen,P.H., Machado,A.C., Laird-Offringa,I.A., Rohs,R. and Chen,L. (2012) DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep.*, **2**, 1197–1206.
37. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Echols,N., Headd,J.J., Hung,L.W., Jain,S., Kapral,G.J., Grosse Kunstleve,R.W. *et al.* (2011) The Phenix software for automated determination of macromolecular structures. *Methods*, **55**, 94–106.
38. Bramucci,E., Paiardini,A., Bossa,F. and Pascarella,S. (2012) PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinformatics*, **13**(Suppl. 4), S2.
39. Zhou,T., Yang,L., Lu,Y., Dror,I., Dantas Machado,A.C., Ghane,T., Di Felice,R. and Rohs,R. (2013) DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.*, **41**, W56–W62.
40. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
41. Li,J., Jiang,L., Liang,X., Qu,L., Wu,D., Chen,X., Guo,M., Chen,Z., Chen,L. and Chen,Y. (2017) DNA-binding properties of FOXP3 transcription factor. *Acta Biochim. Biophys. Sinica*, **49**, 792–799.
42. Homer,N., Merriman,B. and Nelson,S.F. (2009) BFAST: an alignment tool for large scale genome resequencing. *PLoS One*, **4**, e7767.
43. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
44. Dai,S., Sun,C., Tan,K., Ye,S. and Zhang,R. (2017) Structure of thrombospondin type 3 repeats in bacterial outer membrane protein A reveals its intra-repeat disulfide bond-dependent calcium-binding capability. *Cell Calc.*, **66**, 78–89.
45. Overdier,D.G., Porcella,A. and Costa,R.H. (1994) The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Mol. Cell Biol.*, **14**, 2755–2766.
46. Chen,X., Ji,Z., Webber,A. and Sharrocks,A.D. (2016) Genome-wide binding studies reveal DNA binding specificity mechanisms and functional interplay amongst Forkhead transcription factors. *Nucleic Acids Res.*, **44**, 1566–1578.
47. Spek,C.A., Greengard,J.S., Griffin,J.H., Bertina,R.M. and Reitsma,P.H. (1995) Two mutations in the promoter region of the human protein C gene both cause type I protein C deficiency by disruption of two HNF-3 binding sites. *J. Biol. Chem.*, **270**, 24216–24221.
48. Sheng,W., Rance,M. and Liao,X. (2002) Structure comparison of two conserved HNF-3/fkh proteins HFH-1 and genesis indicates the existence of folding differences in their complexes with a DNA binding sequence. *Biochemistry*, **41**, 3286–3293.
49. Stroud,J.C., Wu,Y., Bates,D.L., Han,A., Nowick,K., Paabo,S., Tong,H. and Chen,L. (2006) Structure of the forkhead domain of FOXP2 bound to DNA. *Structure*, **14**, 159–166.
50. Jin,C., Marsden,I., Chen,X. and Liao,X. (1999) Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. *J. Mol. Biol.*, **289**, 683–690.
51. Tsai,K.L., Huang,C.Y., Chang,C.H., Sun,Y.J., Chuang,W.J. and Hsiao,C.D. (2006) Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. *J. Biol. Chem.*, **281**, 17400–17409.
52. Tsai,K.L., Sun,Y.J., Huang,C.Y., Yang,J.Y., Hung,M.C. and Hsiao,C.D. (2007) Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res.*, **35**, 6984–6994.
53. Chen,Y., Chen,C., Zhang,Z., Liu,C.C., Johnson,M.E., Espinoza,C.A., Edsall,L.E., Ren,B., Zhou,X.J., Grant,S.F. *et al.* (2015) DNA binding by FOXP3 domain-swapped dimer suggests mechanisms of long-range chromosomal interactions. *Nucleic Acids Res.*, **43**, 1268–1282.
54. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
55. Nakagawa,S., Gisselbrecht,S.S., Rogers,J.M., Hartl,D.L. and Bulyk,M.L. (2013) DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 12349–12354.
56. Berry,F.B., Lines,M.A., Oas,J.M., Footz,T., Underhill,D.A., Gage,P.J. and Walter,M.A. (2006) Functional interactions between FOXC1 and PITX2 underlie the sensitivity to FOXC1 gene dose in Axenfeld–Rieger syndrome and anterior segment dysgenesis. *Hum. Mol. Genet.*, **15**, 905–919.
57. Berry,F.B., O'Neill,M.A., Coca-Prados,M. and Walter,M.A. (2005) FOXC1 transcriptional regulatory activity is impaired by PBX1 in a filamin A-Mediated manner. *Mol. Cell. Biol.*, **25**, 1415–1424.