

# Novel fold and wing structure of Forkhead transcription factor facilitate DNA binding

George L. Wang <sup>1,†</sup>, Yibei Jiang <sup>1,†</sup>, Yuying Sun <sup>2,†</sup>, Fariborz Nasertorabi <sup>3,†</sup>,  
Jesse A. Weller <sup>1,4</sup>, Raktim Mitra <sup>1,5</sup>, Alexander Batyuk <sup>6</sup>, Oscar M. Aparicio <sup>2</sup>,  
Vadim Cherezov <sup>3,7</sup>, Remo Rohs <sup>1,4,7,8,9,10,\*</sup>

<sup>1</sup>Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, United States

<sup>2</sup>Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, United States

<sup>3</sup>Structural Biology Center, Bridge Institute, Michelson Center for Convergent Bioscience, University of Southern California, Los Angeles, CA 90089, United States

<sup>4</sup>Department of Physics & Astronomy, University of Southern California, Los Angeles, CA 90089, United States

<sup>5</sup>Present address: Institute for Protein Design, University of Washington, Seattle, WA 98195, United States

<sup>6</sup>Linac Coherent Light Source, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, United States

<sup>7</sup>Department of Chemistry, University of Southern California, Los Angeles, CA 90089, United States

<sup>8</sup>Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, CA 90089, United States

<sup>9</sup>Division of Medical Oncology, Department of Medicine, University of Southern California, Los Angeles, CA 90033, United States

<sup>10</sup>Alfred E. Mann Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, United States

\*To whom correspondence should be addressed. Email: rohs@usc.edu

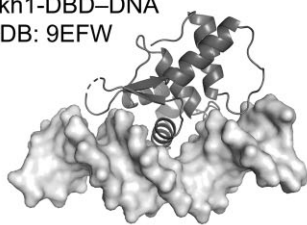
†The first four authors should be regarded as Joint First Authors.

## Abstract

Forkhead homologue 1 (Fkh1) is a yeast transcription factor that plays essential roles in cell-cycle dynamics. Here, we report the co-crystal structure of the DNA-binding domain (DBD) of the yeast Fkh1 protein in complex with a 19-base pair oligonucleotide containing the core binding site and flanking regions. The three-dimensional structure of the Fkh1-DBD reveals a previously unknown protein fold among all known Forkhead proteins. The winged-helix fold forms base-specific contacts of  $\alpha$ -helix H3 with the major groove of the core binding site. Wing 1 and Wing 2 form DNA shape-mediated contacts with the minor groove of the binding site flanking regions. The conformation of Wing 2 is distinct from all known Forkhead proteins, with  $\alpha$ -helices H5 and H6 wrapping back onto the protein core, creating a stable Wing 2 loop. Backbone interactions with  $\beta$ -strands S1 and S2 reveal a structural mechanism for previously observed flanking region preferences in SELEX-seq experiments. *In vivo* yeast experiments on Fkh1 mutants demonstrate that wing residues interacting with flanking regions are important for Fkh1 function. Molecular dynamics simulations relate Fkh1 function to conformational flexibility of wing residues. The novel Forkhead fold enables Fkh1 function with implications, such as structure-based protein design, for other DNA-binding proteins.

## Graphical abstract

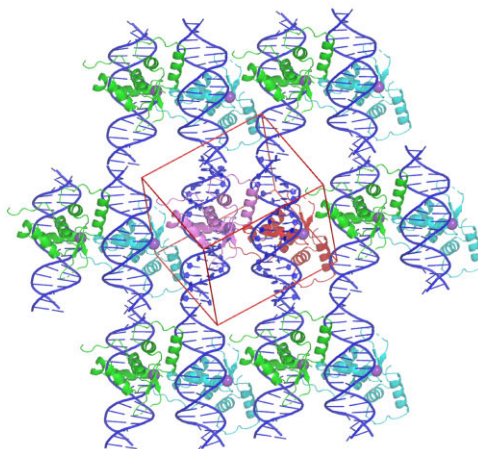
Fkh1-DBD-DNA  
PDB: 9EFW



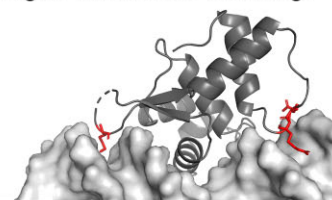
Novel Forkhead fold



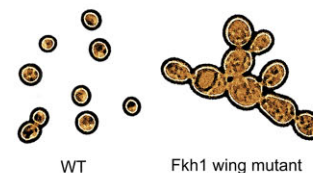
Packing diagram of Fkh1-DBD-DNA crystal structure



Shape-mediated flanking region interactions with wings



Wing mutations result in Fkh1 dysfunction



Received: July 4, 2025. Revised: August 21, 2025. Accepted: September 2, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://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 [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

## Introduction

A member of the Forkhead family of transcription factors (TFs), Forkhead homologue 1 (Fkh1) in *Saccharomyces cerevisiae*, was first identified for its role in silencing mating-type cassette homothallism right a (HMRa) and regulating yeast cell division [1–4]. Its functions involve control of cell cycle-regulated genes [2], stimulation of DNA replication origin firing [5, 6], spatial organization of chromosomes, and regulation of mating-type switching. Other yeast Forkhead proteins include Fkh2, Hcm1, and the more distantly related Fhl1 [7]. Forkhead box (FOX) proteins are a large family of TFs including more than 170 known variants in different species with 19 + subfamilies [8]. They are defined by the presence of a highly conserved winged-helix DNA-binding domain (DBD), commonly known as FKH or Forkhead domain. Differences in DBDs contribute to the varying DNA-binding specificities among FOX proteins [9–11]. Forkhead DBDs are generally ~80–100 amino acids (aa) long and commonly consist of 3–4  $\alpha$ -helices, 2  $\beta$ -strands, and 1–2 winged loops at the C-terminus [8].

FKH–DNA recognition is primarily mediated by the third  $\alpha$ -helix (H3) that forms contacts with the major groove of the bound DNA. These protein–DNA contacts correspond to core binding sites that have the canonical Forkhead DNA motif 5'-RYAAACA-3' (R = A/G, Y = C/T) [9]. Fkh1 binds to one of these canonical motifs, demonstrating its highest affinity for the binding site 5'-GTAAACA-3' [12]. Some FKH domains contact flanking regions adjacent to the core binding site. The minor groove (MG) in these regions displays additional recognition sites that interact with winged loops of the protein, providing enhanced binding affinity and specificity. Previous SELEX-seq experiments showed that flanking nucleotides up to position –4 at the 5' end and +2 at the 3' end of the core binding site have considerable impact on Fkh1 binding [13]. These flanking regions, as well as contacts such as hydrogen bonds, may be sensitive to DNA shape, as opposed to strictly sensitive to sequence, due to electrostatic interactions [13–15].

Although several FOX DBD structures with DNA targets have been reported for other species, to date none has been reported for yeast. Here, we report a 2.2 Å co-crystal structure of the yeast Fkh1 DBD (Fkh1-DBD) (residues 299–421) bound to a 19-base pair (bp) DNA sequence containing the canonical Forkhead binding sequence 5'-GTAAACA-3', with inclusion of high-affinity flanking regions, to investigate the MG contacts of the winged loop. MG contacts are significant due to the documented DNA shape preferences observed in numerous Forkhead proteins [13, 16–19]. This complex provides structural evidence for the influence of the flanking regions of Fkh1 binding sites and reveals a well-structured Wing 2 region that is not present in other FOX proteins. To investigate the DNA-binding mechanism of Fkh1, we generated multiple Fkh1–DNA constructs and conducted molecular dynamics (MD) simulations and BioEmu [20] analyses of the conformational stability and flexibility of these complexes. We identified important residues involved in interactions with flanking regions and introduced mutations in these residues into yeast to determine their effect on Fkh1 function. Our study focused on identifying critical residues within the Wing 2 region of Fkh1 that facilitate DNA binding and explored the binding rules governing sequences of varying binding affinity.

## Materials and methods

### Protein expression and purification

The codon-optimized gene from the DBD (aa 299–421) of FKH1 Yeast Forkhead protein homolog 1 (UniProt ID P40466) was ordered from GenScript, USA. Homology evidence and AlphaFold 2 (AF2) [21] predictions were used to determine the construct length. The gene was inserted in a pET-28a vector with an 8 $\times$ His-tag, followed by a TEV cleavage site on its N-terminus. The expression vector was transformed into T7 Express Competent *Escherichia coli* cells (Cat # C2566, New England Biolabs, USA) for protein expression. A single colony of transformed cells was initially inoculated into 50 ml Luria-Bertani medium (Cat # BP1426-2, Fisher Bioreagent, USA) supplemented with 50  $\mu$ g/ml kanamycin and incubated overnight at 37°C with shaking at 220 RPM. The next day, 20 ml of overnight culture was used to inoculate 2 l of Terrific Broth medium. Inoculated cell culture was grown at 37°C with shaking at 220 RPM until optical density (OD) reached 0.5. Temperature was dropped to 18°C, and the culture was supplied with 0.25 mM IPTG to induce expression. Cells were grown for 16 h after induction and then collected and stored at –80°C until use.

Cells were thawed and dispersed in 100 ml of lysis buffer [50 mM HEPES, pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM AEBSF protease inhibitor, and 1000 units of DENARASE endonuclease (c-LEcta, Germany)] and subjected to three rounds of sonication (time: 3 min, pulse: 3 s, pause: 6 s, amplitude: 85%) using a Q500 sonicator (QSONICA, USA) with a 0.5" standard probe in an ice bath. Lysed cells were centrifuged at 50 000  $\times$  g for 50 min at 4°C to separate cell debris from lysate.

Supernatant was mixed with 1 ml of Talon beads (Takara) pre-equilibrated with lysis buffer and placed on a rotator at 4°C for 30 min. Beads were transferred to a gravity column and thoroughly washed with lysis buffer. Beads underwent an additional wash with 5 ml of low-salt buffer (25 mM HEPES, pH 7.5, 250 mM NaCl, and 10% glycerol) before being eluted with 7 ml of elution buffer (25 mM HEPES, pH 7.5, 250 mM NaCl, 300 mM imidazole, and 10% glycerol). Eluted sample was loaded onto a 1 ml cation exchange column (Hi-Trap SP, Cytiva, USA) and eluted with a gradient against high-salt buffer (25 mM HEPES, pH 7.5, 1 M NaCl, and 10% glycerol). Best fractions by sodium dodecyl sulfate gel were pooled and subjected to overnight cleavage with TEV at 1:50 (molar ratio) TEV:protein at 4°C. 6 $\times$ His TEV was expressed in *E. coli* and purified at the USC Structural Biology Center. Cleaved sample was passed over 0.5 ml of Talon beads pre-equilibrated with low-salt buffer. Flow-through was collected, concentrated, and loaded onto a Superdex 200 Increase 10/300 GL column (Cytiva, USA) pre-equilibrated with 10 mM HEPES, pH 7.5, 200 mM NaCl, and 10% glycerol. Peak fractions corresponding to 1:1 complex were pooled and concentrated to 5 mg/ml using a 5 kDa cutoff centrifugal filter (Sartorius, Germany) in preparation for protein–DNA complex reconstitution.

### Duplex DNA preparation

Single-stranded DNA oligonucleotides were procured from Integrated DNA Technologies (IDT, USA), including 1  $\mu$ mol of the binding sequence and 1  $\mu$ mol of its reverse complement (Supplementary Table S1). IDT formed the duplex of

these two sequences and purified the duplex through non-denaturing polyacrylamide gel electrophoresis (PAGE). Final yield was 142 nmol of purified duplex DNA. This sequence was designed specifically to contain a single Fkh1 binding site 5'-GTAAACA-3' in the center with flanking regions that increase binding affinity while reducing the likelihood of generating multiple binding sites, which could potentially introduce inhomogeneity to the protein–DNA complex sample resulting in impairment of the crystal packing.

### Protein–DNA complex formation

The protein sample was mixed with double-stranded DNA (dsDNA) in a 5:1 molar ratio of protein:DNA. The complex was kept on ice for 30 min before loading onto a Superdex 200 Increase 10/300 GL column pre-equilibrated with 10 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl<sub>2</sub>, and 10% glycerol. Peak fractions corresponding to complex formation were concentrated to 7.5 mg/ml for crystallization using a 5 kDa cutoff centrifugal filter (Sartorius, Germany).

### Protein–DNA co-crystallization

Crystallization was performed at 22°C, using the hanging drop vapor diffusion method in 24-well VDX plates (Hampton Research, USA) with a 1:1 ratio of 1 µl of reservoir solution and 1 µl of protein:DNA solution against 500 µl of reservoir solution. The optimized condition for crystal growth was 100 mM HEPES, pH 7.0, 37.5%–40% PEG 5K-MME, and 200 mM of ammonium phosphate monobasic, which produced small cubic-shaped crystals. Crystals grew to their maximum size of 20 × 30 µm within 4 days, after which they were carefully harvested and flash-frozen in liquid nitrogen and transported to a synchrotron beamline for data collection.

### X-ray diffraction, data collection, and structure determination

Crystallographic data were collected at the Stanford Synchrotron Radiation Lightsources (SSRL) using beamline 12-1 equipped with an Eiger X 16M detector (Dectris). The X-ray beam was attenuated by 80%, and diffraction images were collected from a single crystal using 0.2 s exposure and 0.35° oscillation for a total of 275° of rotation. Collected data were indexed, integrated, and merged with X-ray Detector Software [22] at 2.1 Å resolution, based on the criterion of CC1/2 > 0.5 in the highest resolution shell. The space group was determined as *P1* with two complexes per asymmetric unit and a Matthews coefficient of 2.34. Initial phase information was obtained by molecular replacement using PHASER [23] with an AF2 [21] model of Fkh1 as the initial search model. The structure was improved by iterative rounds of model building and refinement using the programs Coot (v. 0.9.8.92) [24] and the Phenix.refine module of the Phenix software package (v. 1.21–5207) [25, 26]. The dsDNA sequences were added to the model, and model building and refinement continued. The refinement strategy included refining XYZ coordinates in reciprocal space, occupancies, individual B-factors, and TLS groups (one group per chain), as well as using riding hydrogens and NCS restraints. During the final round of refinement, resolution was limited to 2.2 Å because of a slight data anisotropy, NCS restraints were omitted, and the option to optimize the X-ray/stereochemistry weight was

selected. Crystallographic details and statistics are provided in [Supplementary Table S2](#).

### Molecular dynamics simulations

First, we simulated the two complexes observed in the co-crystal structures of the Fkh1–DBD–DNA complex. Next, we performed mutation simulations for the R400 and/or R401 residues in the Fkh1 protein with the sequences seen in the co-crystal structure to test the individual and combined influence of these arginine residues on Fkh1 binding to DNA. Single-residue mutations of the Fkh1 protein were generated using PyMol [27]. To assess the influence of flanking sequences on binding, we varied 4-bp flanking sequences for the 5' end and 2 bp for the 3' end using data from SELEX-seq experiments [13] using Top-Down Crawl [28] for alignment. Four sequences with high, medium-high, medium-low, and low binding affinities were selected ([Supplementary Table S3](#)).

The MD simulation protocol was previously described [29] (for details see [Supplementary Section S-II](#)). We used the near-final refinement version of the Fkh1 protein, which has a root-mean-square deviation (RMSD) of 0.17 Å from the final structure for protein heavy atoms in both Complex 1 and Complex 2. To fill in missing residues in the Fkh1 protein structure (e.g. those in the Wing 1 loop region), we used MODELLER [30] to build complete protein models. To simulate both complexes from the co-crystal structure, we extended the DNA on both ends by 6 bp of GC repeats to achieve a longer, 27-bp sequence. We used Deep DNASHape [31] to predict the DNA shape of the extended DNA sequence. Then, to ensure that the DNA shape of the bound conformation was captured, we replaced the DNA shape of the new sequence to match that from the co-crystal structure.

To construct the DNA used in simulations of sequences from SELEX-seq experiments, we predicted DNA structural characteristics for the flanking sequences with 6-bp GC repeats ([Supplementary Table S3](#)) using the Deep DNASHape webserver [32]. Next, we matched the DNA shape in the core motif of the generated DNA structures to the shape in the co-crystal structure. This step maintains the bound DNA shape and preserves the protein–DNA interactions captured in the co-crystal structure. Finally, the geometry of all constructed DNA oligos was refined and minimized using PHENIX [25] with the *geometry\_minimization* program to avoid any potential steric atom clashes. RMSF values were calculated using GROMACS 2020.3 *rmsf* for C<sub>α</sub> atoms of protein residues for the last 100 ns of the MD trajectory. Additional details are provided in [Supplementary Section S-II](#) and [Supplementary Fig. S1](#) [33–38].

### Biophysical calculations

#### Hydrogen bond analysis

Hydrogen bonds were calculated with the GROMACS 2020.3 [37, 38] *hbond* package. Hydrogen bonds were defined as a distance cutoff of 3.5 Å and a bond angle cutoff of 120° between donor and acceptor. All hydrogen bonds were analyzed for the final 100 ns of simulations.

#### DNA shape analysis

Minor groove width (MGW) and DNA curvature were calculated with Curves 5.3 [39], as previously described [40], using MD snapshots obtained every 100 ps. MGW calculations were plotted for the last 100 ns of simulations.

## Deep-learning studies of protein–DNA binding

### Protein–DNA binding specificity predictions

Structures of protein–DNA complexes were analyzed with DeepPBS [41]. DeepPBS predicts the position weight matrix associated with the DNA binding specificity based on a given conformation of the complex based on geometric deep learning. In this process, DeepPBS computes protein-heavy atom-level relative importance scores (RI scores), which can be aggregated at the residue level. This represents importance scores of protein residues toward determining DNA binding specificity. The DeepPBS webserver [41] was used for this analysis, with the ‘both readout’ model and the option to calculate heavy atom relative importance scores.

### Protein conformational flexibility analysis

Structural variability was probed based on multiple protein conformers generated with Biomolecular Emulator (BioEmu) [20], an approach that samples conformational flexibility in the vicinity of a protein’s equilibrium structure through generative modeling. For ensemble generation, BioEmu [20] was run with protein sequence as input and standard settings.

## Yeast *in vivo* functional testing

### Strains and plasmids

Wild-type (WT) *FKH1* was cloned from the endogenous *FKH1* locus into the XmaI-digested pRS415 plasmid backbone using Gibson Assembly (Gibson Assembly Ultra Kit from Telesis Bio, USA). The forward and reverse primers for amplification are listed in [Supplementary Table S1](#). Mutations were introduced into this plasmid using the QuickChange Lightning Multi Site-directed mutagenesis kit (Agilent 210515 from Agilent, USA). Plasmids were transformed into *fkh1Δ::URA3Mx fkb2Δ::HIS3Mx* strain OAy1121 using the LiAc-PEG method [42] with selection on complete synthetic medium (CSM) minus leucine (Sunrise Science, USA). OAy1121 was a *MATα* version of strain OAy1123 described previously [43]. The pRS415 plasmid includes LEU2, which encodes for a leucine-synthesizing enzyme, so that the yeast strain, which was initially leucine auxotroph, can synthesize leucine upon plasmid transformation. The strain harboring pRS415 was grown in leucine dropout medium to continuously select for cells containing pRS415.

### Microscope images and quantification

Yeast harboring the earlier plasmids were cultured in liquid CSM-leu at 30°C to mid-log phase (OD600 = ~0.7). Next, 500 μl of cell cultures were taken and briefly sonicated using a micro horn (2 pulses, 1 s each with 25% amplitude) before imaging (Branson SFX150). Brightfield images were captured using a DeltaVision Elite Microscope (GE HealthCare, USA) using a 60×/1.42 oil immersion objective lens. For each strain, 25-panel (5 × 5) images were taken and stitched together. Three 25-panel images were quantified. Cells in the field were counted and classified into clusters (>2 cells connected to each other) and non-clusters. The total number of cells in clusters was summed. Percentage of cells in clusters was calculated as the total number of cells in clusters divided by the total number of cells in the field. All images were adjusted for brightness and contrast using ImageJ software.

## Results

### Co-crystal structure of Fkh1-DBD–DNA complex

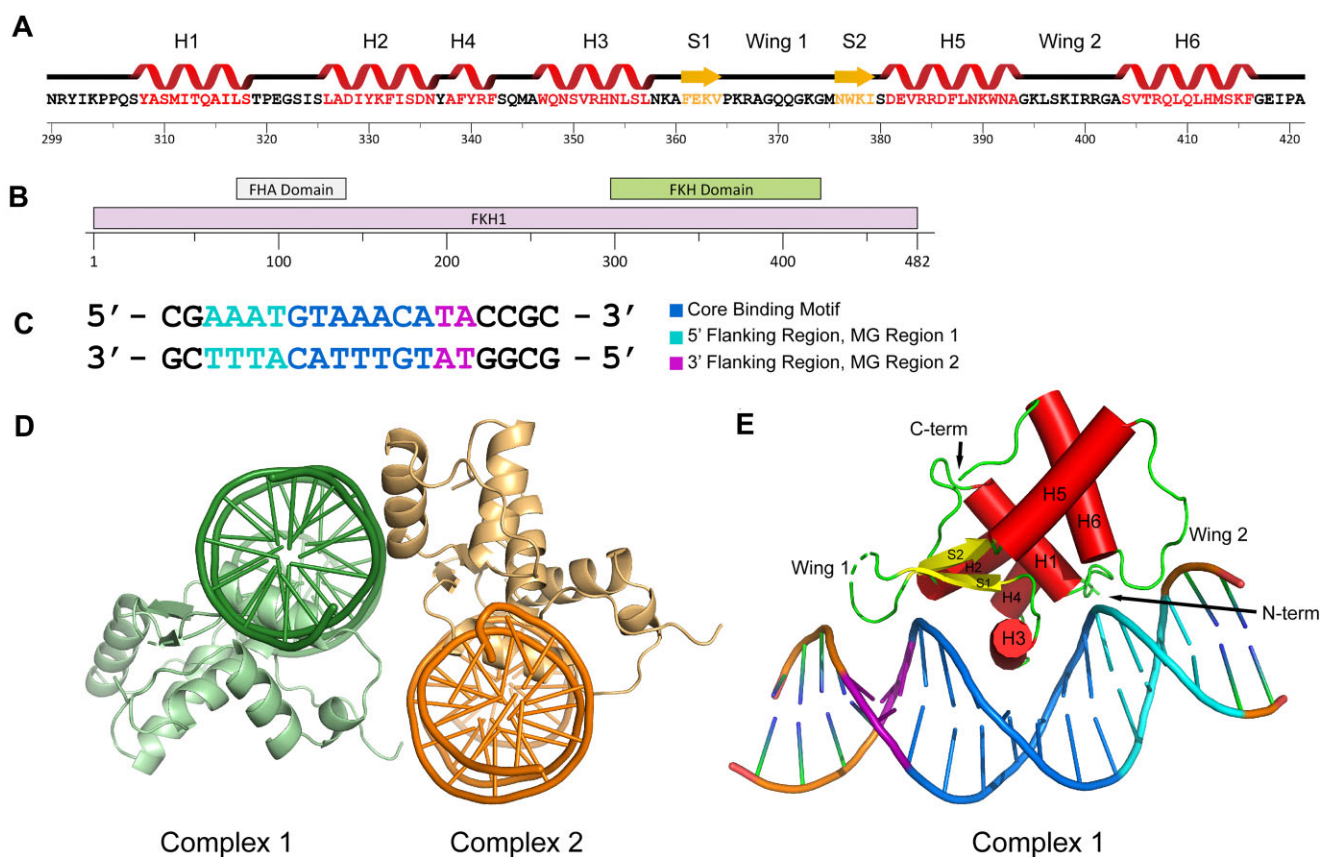
We determined the co-crystal structure of the Fkh1-DBD (aa 299–421) bound to a 19-bp dsDNA containing the canonical binding sequence 5′-GTAAACA-3′. Figure 1 shows a secondary structure plot for the Fkh1-DBD (Fig. 1A), a linear protein map showing the domains of the full-length Fkh1 (Fig. 1B), and the sequence of the bound DNA (Fig. 1C). Crystallographic data were collected from a single crystal that diffracted to a resolution of 2.1 Å and belonged to space group *P1* ([Supplementary Table S2](#)). The solved structure contains two Fkh1-DBD–DNA complexes per asymmetric unit (Fig. 1D). All residues were resolved except for residues 367–371 in both complexes in the asymmetric units, indicating a flexible part of the loop of Wing 1 (Fig. 1D).

### Novel protein fold of Fkh1-DBD

Like other FOX family members, the Fkh1-DBD monomer in Fkh1-DBD–DNA Complex 1 adopts a typical winged-helix structure (Fig. 1E). However, unlike many FOX structures that only have 3–4 α-helices, Fkh1 comprises five α-helices (H1, H2, H3, H5, and H6), one 3<sub>10</sub>-helix (H4), two antiparallel β-strands (S1 and S2), and two winged loops (Wing 1 and Wing 2). The core structure of Fkh1 consists of three stacked α-helices (H1, H2, and H3), a 3<sub>10</sub>-helix (H4) between H2 and H3, and two antiparallel β-strands (S1 and S2) with a loop (Wing 1) between S1 and S2. The structure continues with the stacked α-helices H5 and H6. Helix H6 wraps back into the core structure, stabilizing a loop (Wing 2) that resides between H5 and H6. The structure is further stabilized by the presence of a K<sup>+</sup> cation connecting H3 and S1, as observed in other Forkhead structures [18]. These interactions are visualized in [Supplementary Fig. S2](#). There are minor differences between the two complexes in the asymmetric unit, which seem to be a result of crystal packing (see [Supplementary Section S-III](#) and [Supplementary Figs S3–S11](#)).

Fkh1-DBD is overall longer than most FOX proteins (~120 residues versus ~80–100 residues). Defining the Wing 2 region are the residues beyond the S2 β-strand, resulting in Fkh1 featuring an extended Wing 2 region (~40 residues long). In comparison, many FOX proteins either lack a Wing 2 region entirely or have one that is only ~20 residues long. A few other FOX proteins contain a C-terminal Wing 2 region that is well characterized, and even fewer are associated with DNA interactions (Fig. 2 and [Supplementary Fig. S12](#)). The Fkh1 Wing 2 region features a closed loop that forms base-specific interactions with DNA. The H5 and H6 α-helices stabilize Wing 2 by interacting with the core Forkhead domain, creating a unique Forkhead domain fold that forms well-characterized MG interactions.

Compared to other solved FOX-DBD–DNA structures, the only structure with a well-structured Wing 2 of comparable length is FOXH1. However, whereas the FOXH1 Wing 2 interacts with DNA, it only contacts the DNA backbone and does not form base-specific contacts. Most FOX proteins that form Wing 2 contacts with DNA do so with the DNA backbone. In some cases (e.g. FOXO3a), there are base-specific contacts, but these are only within an unstructured loop. Of the existing FOX-DBD–DNA structures, Fkh1 is the only Forkhead protein to have a well-structured Wing 2 that is known to also form base-specific contacts with the MG.



**Figure 1.** Co-crystal structure of Fkh1-DBD–DNA complex. **(A)** Secondary structure and sequence visualization of Fkh1-DBD (aa 299–421). **(B)** Linear protein map of full-length Fkh1, containing two main domains: FHA domain (aa 76–142) and FKH DBD (aa 299–421). **(C)** Sequence of bound 19-bp DNA. Core binding site and adjacent flanking regions are labeled. **(D)** Overall structure of asymmetric unit cell of Fkh1-DBD bound to oligonucleotide 5'-CGAAATGTAAACATACCGC-3'. The asymmetric unit cell contains two separate complexes, Complex 1 and Complex 2, with slightly different conformations. **(E)** Structural representation of Complex 1 in the asymmetric unit, featuring Fkh1-DBD (Chain A) bound to DNA (Chains E and F). Secondary structure elements are labeled as shown for protein sequence in panel (A). Core binding site and flanking regions are marked with the same colors as in panel (C).

### DNA recognition in Fkh1-DBD–DNA complex

As with many FOX-DBD–DNA structures, Fkh1-DBD binds to the core binding site 5'-GTAAACA-3' via interactions between  $\alpha$ -helix H3 of Fkh1-DBD and the DNA major groove, with strong electron density evidence (Supplementary Fig. S13A). The interaction network comprises most of the base-specific contacts of the core binding site. More specifically, N349 forms a bidentate hydrogen bond with the A10 base. R352 forms a hydrogen bond with the G8' base and possibly a stacking interaction with the T7' base. H353 forms hydrogen bonds with the T10' and T11' bases, as well as potential contacts or other hydrophilic interactions with the T8 and A9 bases. S356 forms a hydrogen bond with the phosphate of T9'. The R352, H353, and S356 residues mediate binding to the core binding site in the major groove via conserved hydrogen bonds common to many Forkhead proteins (Fig. 3A). In addition to hydrogen bonds, these residues are major contributors to water-mediated contacts, which can confer another layer of binding specificity (Supplementary Fig. S14).

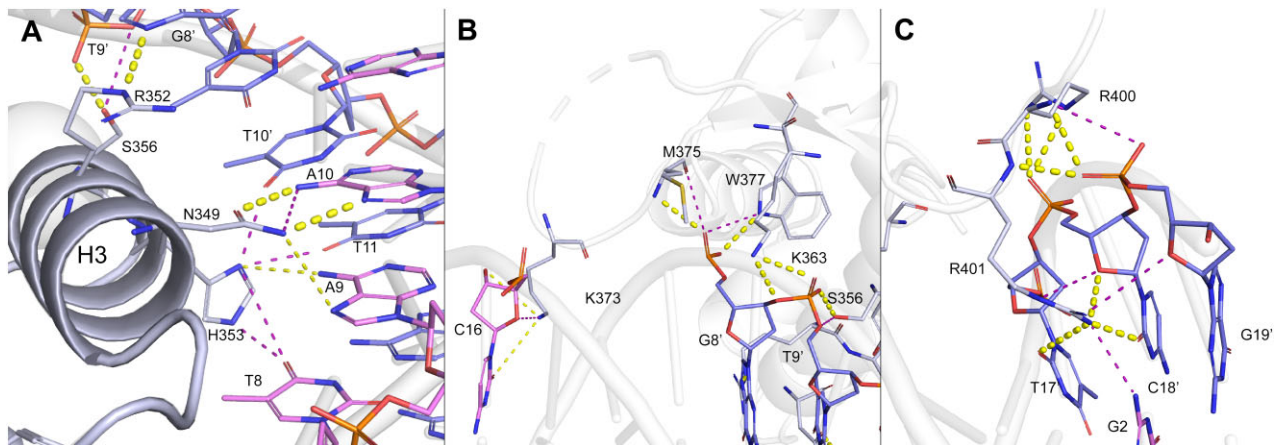
Wing 1 of the Fkh1-DBD interacts with MG region 2 at the 3' flanking region (underlined) of the core binding site 5'-GTAAACATA-3'. The most likely rotamer for the K373 side chain interacts with the sugar and phosphate moieties of C16, although this residue exhibits poor electron density, suggesting conformational flexibility when binding to DNA

(Supplementary Fig. S13C and D). Residue M375 forms a hydrogen bond between its backbone nitrogen and the phosphate of G8'. These two residues also contribute to water-mediated hydrogen bonds with DNA, which are potentially important for differentiating sequences from high to low binding affinity (Supplementary Fig. S14). However, these weaker interactions align with knowledge that the 3' flanking region is shorter and contributes less to Fkh1 binding affinity than the 5' flanking region [13]. These interactions are visualized in Fig. 3B.

Wing 2 of the Fkh1-DBD interacts with MG region 1 at the 5' flanking region (underlined) of the core binding site 5'-AAATGTAAACA-3' and is well characterized with strong electron density (Supplementary Fig. S13B). Such interaction is uncommon for most FOX-DBD structures, which typically have less secondary structure accompanying Wing 2. The Wing 2 interactions occur mainly via two residues, R400 and R401. The sidechain of R400 forms a salt bridge with the phosphate of C18' and a stacking interaction with R407. The backbone nitrogen of R400 forms a hydrogen bond with the phosphate of G19'. The sidechain of R401 forms strong hydrogen bonds with the G2 and C18' bases and the sugar of C18'. The backbone nitrogen of R401 forms a hydrogen bond with the phosphate of G19' (Fig. 3C). R400 and R401 contribute to a highly complex water-mediated hydrogen bond

Name/PDB ID	Secondary Structure Map	Length (bp)	DBD Length (aa)	Wing 2 Length (aa)	Wing 2 Exists (~10aa)	Wing 2 is Structured	Wing 2-DNA Backbone Contact (< 3.5 Å)	Wing 2-DNA Base-Specific Contact (< 3.5 Å)	Species
Fkh1 (9EFW)		19	123	42	✓	✓	✓	✓	Yeast
FOXA1 (1VTN)		13	102	23	✓	✗	✗	✗	Human
FOXA2 (7YZE)		16	87	3	✗	✗	✗	✗	Human
FOXC2 (6LBM)		20	103	24	✓	✗	✓	✗	Human
FOXG1 (7CBY)		16	100	19	✓	✓	✓	✗	Human
FOXH1 (7YZB)		16	132	47	✓	✓	✓	✗	Human
FOXL2 (7VOU)		16	85	8	✗	✗	✗	✗	Human
FOXN1 (6EL8)		13	100	18	✓	✓	✓	✗	Human
FOXN3 (6NCE)		16	100	19	✓	✓	✓	✗	Human
FOXO1 (3CO7)		13	87	2	✗	✗	✗	✗	Human
FOXO3a (2UZK)		13	97	17	✓	✗	✓	✓	Human
FOXO4 (3L2C)		13	85	1	✗	✗	✗	✗	Human
FOXFP3 (7TDX)		14	90	4	✗	✗	✗	✗	Mouse

**Figure 2.** Table detailing key similarities and differences between the Fkh1-DBD-DNA complex and other available Forkhead protein-DNA structures. The comparison reveals multiple novelties of the Fkh1-DBD-DNA structure that have previously not been observed. It exhibits a novel protein fold, the longest structured Wing 2 region that has been observed, and one of the longest DNA fragments bound to a Forkhead protein.



**Figure 3.** DNA recognition of Fkh1-DBD domain. Hydrogen bonds are shown in yellow (distance < 3.5 Å). Less favorable polar interactions are shown in purple (distance < 4.0 Å). (A) Interactions between  $\alpha$ -helix H3 and major groove of bound DNA. (B) Interactions between K373 of Wing 1 and MG region 2, as well as interactions between K363 and W377 of  $\beta$ -strands S1 and S2 with the DNA backbone. (C) Interactions between R400 and R401 on Wing 2 and MG region 1.

network, conferring binding specificity by the Wing 2 region (Supplementary Fig. S14).

There are additional backbone interactions, including an interaction between  $\alpha$ -helix H1 and the core binding site, where the sidechain of Y308 forms a hydrogen bond with the phosphate of T8, while the backbone nitrogen of Y308 forms a hydrogen bond with the phosphate of G7.  $\beta$ -strands S1 and

S2 interact with the 3' flanking region alongside Wing 1. In  $\beta$ -strand S1, K363 forms a strong salt bridge with the phosphate of T9'. In  $\beta$ -strand S2, W377 forms a hydrogen bond with the phosphate of G8' (Supplementary Fig. S3A).

Overall, there are protein-DNA interactions across the core binding site extending slightly beyond the flanking regions. These results provide structural evidence for the extended

DNA binding motif of Fkh1 whose increased relative binding affinity was observed in SELEX-seq experiments [13].

### R401 residue modulates interaction with minor groove

The Wing 2–DNA interaction is of interest due to its function in achieving binding specificity [18, 43–47]. The Wing 2 domain is highly dynamic and hard to capture in crystal structures [18]. A previous study examined binding of the Wing 2 domain in human, frog, and fish Forkhead proteins [48]. Here, we observe a stronger interaction of the Wing 2 region of yeast Fkh1 with DNA in the co-crystal structure, highlighting the potential significance of residues R400 and R401. These interactions are shown in [Supplementary Fig. S3A](#), visualized using DNAProDB [49].

Among various other Forkhead proteins, multiple consecutive arginine or lysine residues are common in the Wing 2 region. We attempted to analyze the binding pose of several other Forkhead proteins using AlphaFold 3 (AF3) [50]. Predicted structures were aligned by the first 80 residues (covering the core and Wing 1 regions) to analyze differences in the Wing 2 regions. Predicted poses of two arginine residues, aligned with R400 and R401 in our co-crystal structure, are shown as sticks in [Supplementary Fig. S15](#). Most predicted arginine residues interact with the DNA backbone rather than with MG region 2. With a relatively low average AF3 confidence score (*pLDDT*) of 66.9 for the two arginine residues, the accuracy of AF3 predictions in this region remains unclear. Meanwhile, the predicted pose for the highly conserved lysine residue, K373, is highly similar among Forkhead proteins.

We also used AF3 [50] to predict the Fkh1-DBD–DNA complex using the same nucleotide and amino acid sequences observed in the co-crystal structure. Overall, predicted structures closely resemble the co-crystal structure ([Supplementary Fig. S16A](#)) with an all-atom RMSD of 2.19 Å for the entire complex. Components of the complex are predicted with high confidence, with an average overall *pLDDT* of 81.31 for the DNA and 88.83 for the protein chain. Predictions for the K373 residue show a relatively high *pLDDT* of 73.51 ([Supplementary Fig. S16B](#), left panel). Despite the poor electron density observed for K373 in the co-crystal structure, the AF3 prediction aligns with this pose, confirming the accuracy of K373 positioning in the co-crystal structure ([Supplementary Fig. S13C](#) and [D](#)). However, predictions for arginine residues R400 and R401 show low average *pLDDT* of 58.51 and 57.71, respectively, among the top five predicted structures ([Supplementary Fig. S16B](#)). Three models predict R401 to be outside MG region 2 (similar to Complex 2 in our co-crystal structure), while two models predict a conformation of R401 similar to that observed in Complex 1 of the co-crystal structure ([Supplementary Fig. S16B](#), right panel). However, none of the predicted orientations for R400 align with its position in the co-crystal structure ([Supplementary Fig. S16B](#), center panel). Despite recent advances in protein structure predictions [51], accurate predictions at atomistic resolution crucial for protein–DNA interactions remain challenging. Therefore, co-crystal structures remain essential for analyzing mechanistic aspects of protein–DNA interactions.

Next, we analyzed the R400 and R401 interactions with the DNA in mechanistic detail. Initially, we ran MD simulations for Complex 1 and generated a mean structure by clustering the trajectory with an RMSD cutoff of 1 Å. This mean

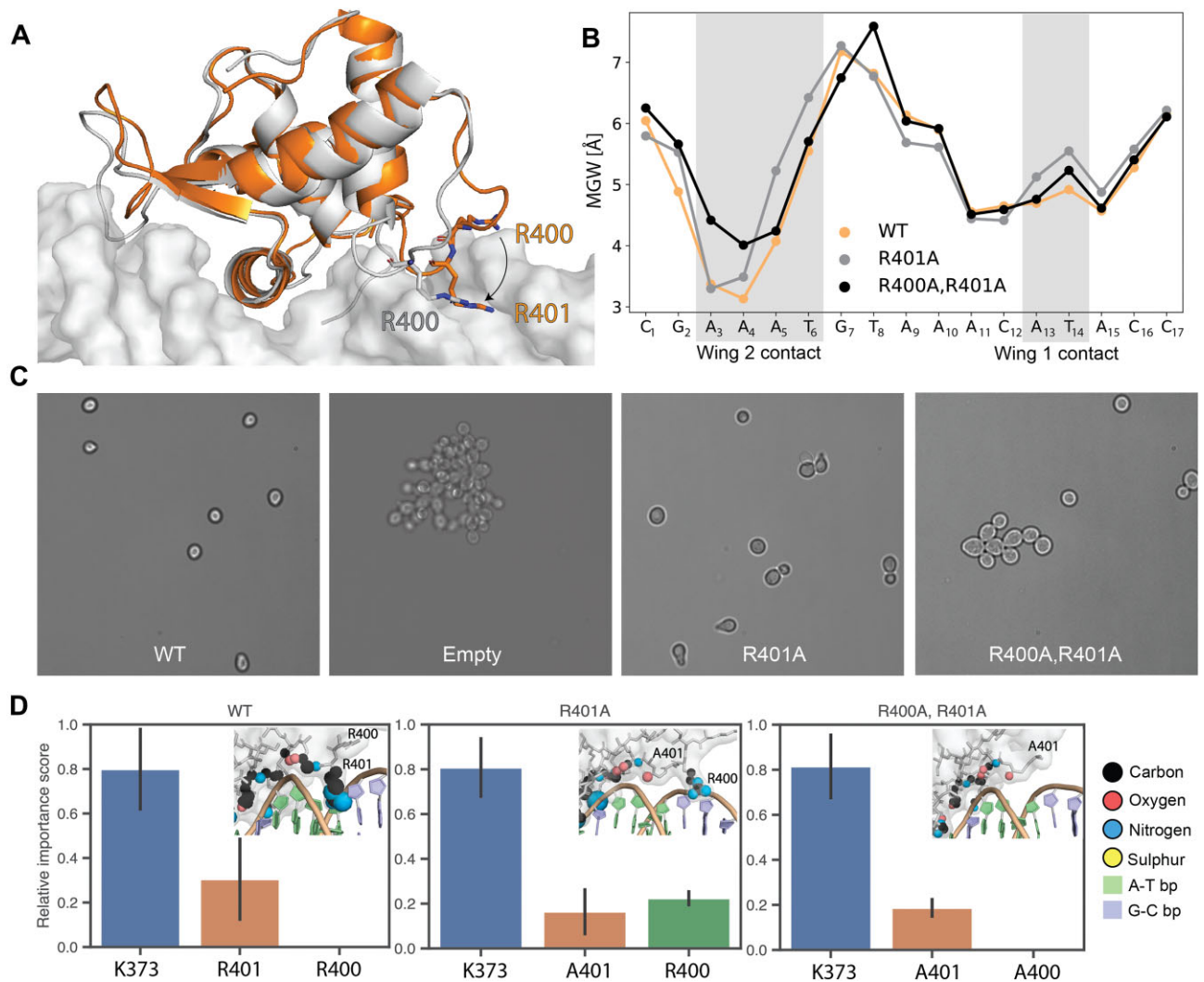
structure was further analyzed with DeepPBS [41], a geometric deep learning-based method that assesses the importance score of each residue for DNA binding specificity by neural network perturbation. In the Wing 2 region, both R400 and R401 exhibit high importance, whereas K373 is most critical in the Wing 1 region. In the core region, residues N349, R352, S356, and H353 are crucial for DNA binding, serving as a control ([Supplementary Fig. S17](#)). We investigated interaction of the R401 residue with DNA by analyzing the correlation between their distance and MGW. We observed a high negative correlation of  $-0.66$  between movements of R401 and MGW ([Supplementary Fig. S18](#)). This finding shows that R401, through electrostatic interactions, can modulate MGW at the flanking region contacting Wing 2. Similar correlations were observed across other replicas with an average absolute correlation of 0.55.

### Mutations reveal substitution mechanism of R400 and R401 in DNA recognition

We analyzed the effect of mutating R401 to alanine (R401A) on DNA binding by conducting MD simulations of the WT and mutated complexes for 300 ns. Clustering the trajectory to derive a mean structure reveals that in the absence of R401, R400 can compensate by interacting with MG region 2 ([Fig. 4A](#)). Specifically, in the R401A mutant, R400 influences the MGW at the A4 base ([Fig. 4B](#) and [Supplementary Fig. S19A](#)) in place of R401. We further quantified this substitution by measuring the distance between the guanidinium group of R400 and the N3 atom of the A4 base. In WT protein, the R400 residue is located at a distance of  $\sim 12.8$  Å from the A4 nucleotide. This distance decreases to  $\sim 5.7$  Å in the R401A mutant, where R400 substitutes for R401 in contacting the MG ([Supplementary Fig. S19A](#)). We calculated these distances for both R401A and WT simulations and found such substitution in most replicas ([Supplementary Fig. S19B](#)).

### Functional validation of Fkh1 wing–DNA contacts and arginine substitution mechanism

Fkh1 and Fkh2 are crucial for regulation of CLB2 cluster genes. To investigate the importance of the R400 and R401 residues for interactions with the DNA flanking region contacted by Wing 2, we tested mutant alleles of *FKH1* for their ability to complement for loss of *FKH1* function. Yeast with both *FKH1* and *FKH2* deleted (*fkh1Δ fkh2Δ*) exhibit pseudohyphal growth [2], in which the yeast form branched structures of connected elongated cells. Cells expressing either *FKH1* or *FKH2* exhibit comparatively normal cell division and morphology, indicating that either TF can compensate for the absence of the other [1–4]. When we introduced WT *FKH1* and its mutant alleles on a plasmid into *fkh1Δ fkh2Δ* cells and visualized cellular morphology ([Fig. 4C](#)), we observed full rescue of the mutant phenotype upon expression of WT *FKH1*. Expression of either the R400A or R401A allele also rescued the pseudohyphal phenotype. However, expression of the double mutant (R400A, R401A) did not fully complement *FKH1* function, with these cells exhibiting a partial pseudohyphal phenotype compared to *fkh1Δ fkh2Δ* cells, characterized by fewer and smaller cell clusters ([Fig. 4C](#) and [Supplementary Fig. S20](#)). This result provides evidence for the substitution mechanism whereby R400 substitutes for DNA binding in the absence of R401, as indicated by the MD simulation analysis.



**Figure 4.** MD simulations and functional assays reveal importance of R401 in protein–DNA binding. **(A)** Structural alignment of R401A mutant (gray) and WT (orange) complexes from clustered MD simulations. Arrow indicates movement of R400 into MG region 2 in R401A system. Arginine residues are shown in sticks. **(B)** DNA shape comparison of mutants (orange and black) with WT complexes (grey). Regions contacted by winged domains are labeled. **(C)** Microscopic images of *fkh1Δ fkh2Δ* strains containing plasmids expressing indicated *FKH1* alleles or empty vector. **(D)** Residue relative importance scores calculated for AF3-predicted WT and mutants using DeepPBS. Visualization of Wing 2 region interactions of each system is provided.

We additionally examined the *FKH1* residue K373 on Wing 1 that contacts DNA in the co-crystal structure (Fig. 3B). The K373A allele failed to fully rescue pseudohyphal growth, consistent with K373 stabilizing the DNA contact with Wing 1. Finally, we combined the Wing 1 and Wing 2 mutations to create the triple mutant (K373A, R400A, R401A). This allele showed an additive effect of the Wing 1 and Wing 2 mutations, although the triple mutant allele was still able to partially complement *FKH1* function (Supplementary Fig. S21). Taken together, the results suggest that both Fkh1 wings assist in stabilizing protein–DNA interactions *in vivo* and that their loss causes a partial defect of Fkh1 function.

#### Deep learning-based analyses emphasize importance of Wing 2 arginine residues in DNA readout

To further confirm the crucial role of R400 and R401 in achieving DNA binding specificity, we used AF3 [50] to predict the structure of the mutant constructs (Fig. 4D). The R400 substitution in R401A is observed in AF3-predicted struc-

tures. Using DeepPBS [41], we calculated the relative importance scores for these two arginine residues. R400 is predicted to interact with DNA only in the R401A system (Fig. 4D, center panel). The role of residues R400 and R401 in Fkh1–DNA binding is further supported through MD simulation analyses (Supplementary Figs S22–S25 and Supplementary Table S4) and a BioEmu [20] analysis of protein conformational flexibility (Supplementary Fig. S26).

#### Discussion

We determined a 2.2 Å resolution co-crystal structure of the Fkh1–DBD–DNA complex and identified the structural mechanisms by which Fkh1 recognizes its DNA target site. Helix H3 forms major groove interactions with the core binding site, while Wing 1 and Wing 2 interact with the MG of the flanking regions. Wing 2 has a stable conformation that has not been observed in other Forkhead proteins and contributes to binding affinity via DNA shape-mediated interactions. Our experimental and MD analyses provide compelling evidence

for the mechanisms responsible for the DNA-binding specificity of flanking regions beyond the core binding site.

Our previous SELEX-seq experiments indicated DNA flanking sequence preferences of Fkh1 and other yeast Forkhead proteins, implying the presence of protein–DNA interactions involving flanking regions to the core binding site [13], but did not allow a description of structural mechanisms responsible for this readout mode based on experimental data. In our prior study, we used AF2 [21] to predict the Fkh1-DBD structure, with similar findings to the X-ray crystallography results obtained here (Supplementary Fig. S26 in Ref. [13]). The RMSD between predicted and final DBD structures was only 0.47 Å. Notably, AF2 predicted the novel Wing 2 fold observed in the co-crystal structure, despite the current absence of other Forkhead proteins with a similar fold in the Protein Data Bank (PDB) [52].

Here, we employed AF2 for the design of the Fkh1-DBD construct used for crystallization. AF2 was useful for detecting the importance of the H5 and H6 helices for the stable Wing 2 region. Protein construct design for crystallization generally favors closely trimmed constructs, as disordered regions can interfere with the formation of crystal contacts. Therefore, the construct must be chosen in a way that trims enough residues to make it suitable for crystallization while still capturing relevant structural information. The final Fkh1-DBD construct contains 123 residues, longer than most Forkhead domains (typically ~80–100 aa). Had we designed the Fkh1 construct based on homology and secondary structure evidence, parts of the H5 and/or H6 helices would have been excluded, resulting in a crystal structure with missing information or likely failure to crystallize. Thus, AF2 was instrumental for expediting the process of producing the crystal and solving the three-dimensional structure of a complete Fkh1-DBD.

The asymmetric unit of the Fkh1-DBD–DNA co-crystal structure features two protein–DNA complexes with slightly different conformations, along with a few interactions between complexes. Human Forkhead-DBD–DNA complexes have been known to dimerize through protein–protein interactions, creating a bridge between two DNA sequences. In some cases, such as FOXP3 [53, 54], this bridging effect can be formed via a domain swap, whereby structural elements are exchanged with adjacent subunits, creating an intertwined multimer structure. Despite appearing to be a dimer in the asymmetric unit, Fkh1 does not exhibit these properties in solution. As observed from the size-exclusion chromatogram and Native-PAGE gel (Supplementary Figs S7 and S8), Fkh1-DBD–DNA exists only as individual complex in solution. Crystal contact analysis with PISA [55] suggested that individual protein–DNA complexes in the asymmetric unit remain stable in solution, whereas the dimer of two protein–DNA complexes in the asymmetric unit is unstable (Supplementary Figs S9 and S10). The crystal packing (Supplementary Fig. S11) further indicates that the composition of the asymmetric unit includes two protein–DNA complexes due to their slight structural variation.

When preparing protein–DNA complexes for crystallization, it is often desirable to isolate the complex from free protein and free DNA. Yet in the literature, FOX-DBD–DNA complexes are often prepared by mixing protein with DNA at a particular molar ratio, without further processing [48, 53], likely because methods such as size-exclusion chromatography can disrupt Forkhead protein–DNA interactions. Interestingly, the Fkh1-DBD–DNA complex was unusually stable

through the size-exclusion chromatography process. This stability could be due to a longer lifetime or higher thermodynamic stability, potentially owing to wing interactions with the flanking regions of the high-binding-affinity oligo. Nevertheless, further analysis of multiple FOX-DBD–DNA complexes via methods such as isothermal titration calorimetry and competition assays would be necessary to determine the cause of improved stability.

The structure of the Fkh1-DBD Wing 2 region enables it to form MG contacts with the flanking region of the core binding site in a manner that primarily depends on interactions of R400 and R401 with DNA. Our MD simulation analyses of selected sequences with varying binding affinity provided results consistent with these observations. Wing 2 stability could be important for both Fkh1 function and efficient Fkh1 binding to target loci in the yeast genome. Recent studies into the C-terminal region of Fkh1-DBD [56] have shown that deletions involving helix H6 and the region directly following it result in Fkh1 dysfunction in *fkh2Δ* yeast. Deletion of residues in the 417–423 region is associated with impaired Fkh1 function and reduced binding efficiency of target loci, consistent with our observations in the Fkh1-DBD–DNA co-crystal structure. In particular, the I419 and P420 residues on the C-terminal end of the H6 helix form hydrogen bonds and hydrophobic interactions with the H5 helix, seemingly stabilizing Wing 2 (Supplementary Fig. S27).

Fkh1–DNA recognition is primarily determined by the H3 recognition helix, which interacts in a sequence-specific manner with the major groove of the bound DNA. The wings of the FKH domain may interact with DNA flanking regions to enhance the binding specificity and/or avidity of interaction with target DNA sequences. For instance, the mutation of the Wing 1 residue K373A was previously shown to impair Fkh1 function [57]. Indeed, we showed here that the wing interactions with the MG of flanking regions [58] are important for Fkh1 function. This was demonstrated by our MD simulation analysis and *in vivo* experiments, which validated the arginine substitution mechanism whereby R400 can substitute for the DNA binding interactions of R401 if mutated (Fig. 4C and Supplementary Figs S18 and S19). Mutations of important wing residues resulted in a degree of pseudohyphal morphology consistent with Fkh1 dysfunction, and mutations on both wings had an additive effect resulting in enhanced dysfunction. Fkh1 function was measurably and clearly diminished by the loss of wing interactions with the MG in the flanking region, indicating the importance of DNA interactions outside of the core binding site for Fkh1 function.

## Acknowledgements

The authors thank Haim Rozenberg of the Weizmann Institute of Science in Rehovot, Israel, for critical assistance with the structural refinement and analysis of crystal packing. Haim Rozenberg elected to be recognized for his contributions as an acknowledgment. We thank Helen M. Berman of the University of Southern California for valuable advice on the project. We thank all members of the Rohs lab for their helpful suggestions, including Brendon H. Cooper for assistance with SELEX-seq data analysis and Yan Gan for help in establishing Fkh1 purification and expression protocols. We also thank Carolyn Phillips of the University of Southern California for providing access to a DeltaVision Elite Microscope for imaging. Experimental projects in the Rohs lab are

supported through the USC Michelson Center for Convergent Bioscience.

**Author contributions:** George Lee Wang (Conceptualization [equal], Data curation [lead], Formal Analysis [lead], Investigation [lead], Visualization [equal], Writing—original draft [lead], Writing—review & editing [lead]), Yibei Jiang (Conceptualization [equal], Data curation [lead], Formal analysis [lead], Investigation [lead], Visualization [equal], Writing—original draft [lead], Writing—review & editing [equal]), Yuying Sun (Data curation [equal], Formal analysis [supporting], Investigation [equal], Validation [equal], Writing—review & editing [supporting]), Fariborz Naserorabi (Data curation [equal], Formal analysis [supporting], Investigation [equal], Visualization [equal], Writing—review & editing [supporting]), Jesse Weller (Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Validation [supporting]), Raktim Mitra (Methodology [supporting], Resources [supporting], Validation [supporting], Visualization [supporting]), Alexander Batyuk (Data curation [supporting], Investigation [supporting]), Oscar M. Aparicio (Data curation [supporting], Funding acquisition [supporting], Investigation [supporting], Methodology [supporting], Validation [equal], Writing—review & editing [supporting]), Vadim Cherezov (Data curation [supporting], Formal analysis [equal], Funding acquisition [supporting], Investigation [supporting], Visualization [equal], Writing—review & editing [supporting]), Remo Rohs (Conceptualization [lead], Formal analysis [supporting], Funding acquisition [lead], Investigation [supporting], Project administration [lead], Supervision [lead], Writing—original draft [equal], Writing—review & editing [equal]).

## Supplementary data

Supplementary data is available at NAR online.

## Conflict of interest

None declared.

## Funding

This work was supported by the National Institutes of Health [R35GM130376 to R.R.; R01GM065494 to O.M.A.], a University of Southern California Office of Research and Innovation SBIR/STTR Planning Award [to R.R.], and an Andrew J. Viterbi Fellowship in Computational Biology and Bioinformatics [to R.M.]. Use of the Stanford Synchrotron Radiation Light source, SLAC National Accelerator Laboratory, was supported by the U.S. Department of Energy [DE-AC02-76SF00515]. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and National Institutes of Health [P30GM133894]. Funding to pay the Open Access publication charges for this article was provided by National Institute of General Medical Sciences [R35GM130376].

## Data availability

The co-crystal structure coordinates of the Fkh1-DBD-DNA complexes and corresponding structure factors have been deposited in the Protein Data Bank under accession code PDB ID 9EFW. The PDB validation report is included in Sup-

plementary Data. MD simulation protocols are accessible and can be retrieved from Zenodo at <https://doi.org/10.5281/zenodo.14166566>. The *in vivo* microscopy images are available through Figshare at <https://doi.org/10.6084/m9.figshare.27913854>.

## References

- Hollenhorst PC, Bose ME, Mielke MR *et al.* Forkhead genes in transcriptional silencing, cell morphology and the cell cycle: overlapping and distinct functions for FKH1 and FKH2 in *Saccharomyces cerevisiae*. *Biophys J* 2000;154:3765–7. <https://doi.org/10.1093/genetics/154.4.1533>
- Zhu G, Spellman PT, Volpe T *et al.* Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* 2000;406:90–4. <https://doi.org/10.1038/35017581>
- Pic A, Lim FL, Ross SJ *et al.* The forkhead protein Fkh2 is a component of the yeast cell cycle transcription factor SFF. *EMBO J* 2000;19:3750–61. <https://doi.org/10.1093/emboj/19.14.3750>
- Kumar R, Reynolds DM, Shevchenko A *et al.* Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. *Curr Biol* 2000;10:896–906. [https://doi.org/10.1016/S0960-9822\(00\)00618-7](https://doi.org/10.1016/S0960-9822(00)00618-7)
- Knott SRV, Peace JM, Ostrow AZ *et al.* Forkhead transcription factors establish origin timing and long-range clustering in *S. cerevisiae*. *Cell* 2012;148:99–111. <https://doi.org/10.1016/j.cell.2011.12.012>
- Peace JM, Villwock SK, Zeytounian JL *et al.* Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Res* 2016;26:365–75. <https://doi.org/10.1101/gr.196857.115>
- Mazet F, Yu JK, Liberles DA *et al.* Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* 2003;316:79–89. [https://doi.org/10.1016/S0378-1119\(03\)00741-8](https://doi.org/10.1016/S0378-1119(03)00741-8)
- Golson ML, Kaestner KH. Fox transcription factors: from development to disease. *Development* 2016;143:4558–70. <https://doi.org/10.1242/dev.112672>
- Carlsson P, Mahlapuu M. Forkhead transcription factors: key players in development and metabolism. *Dev Biol* 2002;250:1–23. <https://doi.org/10.1006/dbio.2002.0780>
- Cirillo LA, Zaret KS. Specific interactions of the wing domains of FOXA1 transcription factor with DNA. *J Mol Biol* 2007;366:720–4. <https://doi.org/10.1016/j.jmb.2006.11.087>
- Boura E, Silhan J, Herman P *et al.* Both the N-terminal loop and wing W2 of the forkhead domain of transcription factor Foxo4 are important for DNA binding. *J Biol Chem* 2007;282:8265–75. <https://doi.org/10.1074/jbc.M605682200>
- Hollenhorst PC, Pietz G, Fox CA. Mechanisms controlling differential promoter-occupancy by the yeast forkhead proteins Fkh1p and Fkh2p: implications for regulating the cell cycle and differentiation. *Genes Dev* 2001;15:2445–56. <https://doi.org/10.1101/gad.906201>
- Cooper BH, Dantas Machado AC, Gan Y *et al.* DNA binding specificity of all four *Saccharomyces cerevisiae* forkhead transcription factors. *Nucleic Acids Res* 2023;51:5621–33. <https://doi.org/10.1093/nar/gkad372>
- Zhou T, Shen N, Yang L *et al.* Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc Natl Acad Sci USA* 2015;112:4654–9. <https://doi.org/10.1073/pnas.1422023112>
- Rohs R, West SM, Sosinsky A *et al.* The role of DNA shape in protein–DNA recognition. *Nature* 2009;461:1248–53. <https://doi.org/10.1038/nature08473>
- Rogers JM, Waters CT, Seegar TCM *et al.* Bispecific forkhead transcription factor FoxN3 recognizes two distinct motifs with

- different DNA shapes. *Mol Cell* 2019;74:245–53. <https://doi.org/10.1016/j.molcel.2019.01.019>
17. Newman JA, Aitkenhead H, Gavard AE *et al.* The crystal structure of human forkhead box N1 in complex with DNA reveals the structural basis for forkhead box family specificity. *J Biol Chem* 2020;295:2948–58. <https://doi.org/10.1074/jbc.RA119.010365>
  18. Dai S, Qu L, Li J *et al.* Toward a mechanistic understanding of DNA binding by forkhead transcription factors and its perturbation by pathogenic mutations. *Nucleic Acids Res* 2021;49:10235–49. <https://doi.org/10.1093/nar/gkab807>
  19. Ibarra IL, Hollmann NM, Klaus B *et al.* Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nat Commun* 2020;11:124. <https://doi.org/10.1038/s41467-019-13888-7>
  20. Lewis S, Hempel T, Jiménez-Luna J *et al.* Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* 2025;389:eadv9817. <https://doi.org/10.1126/science.adv9817>
  21. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>
  22. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr* 2010;66:125–32. <https://doi.org/10.1107/S0907444909047337>
  23. McCoy AJ, Grosse-Kunstleve RW, Adams PD *et al.* Phaser crystallographic software. *J Appl Crystallogr* 2007;40:658–74. <https://doi.org/10.1107/S0021889807021206>
  24. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–32. <https://doi.org/10.1107/S0907444904019158>
  25. Adams PD, Afonine PV, Bunkóczi G *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010;66:213–21. <https://doi.org/10.1107/S0907444909052925>
  26. Liebschner D, Afonine PV, Baker ML *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 2019;75:861–77. <https://doi.org/10.1107/S2059798319011471>
  27. Schrödinger LLC. The PyMOL molecular graphics system. *Version* 2015;1:8.
  28. Cooper BH, Chiu TP, Rohs R. Top-Down Crawl: a method for the ultra-rapid and motif-free alignment of sequences with associated binding metrics. *Bioinformatics* 2022;38:5121–3. <https://doi.org/10.1093/bioinformatics/btac653>
  29. Jiang Y, Chiu TP, Mitra R *et al.* Probing the role of the protonation state of a minor groove-linker histidine in Exd-Hox–DNA binding. *Biophys J* 2024;123:248–59. <https://doi.org/10.1016/j.bpj.2023.12.013>
  30. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
  31. Li J, Chiu TP, Rohs R. Predicting DNA structure using a deep learning method. *Nat Commun* 2024;15:1243. <https://doi.org/10.1038/s41467-024-45191-5>
  32. Li J, Rohs R. Deep DNASHape webserver: prediction and real-time visualization of DNA shape considering extended *k*-mers. *Nucleic Acids Res* 2024;52:W7–12. <https://doi.org/10.1093/nar/gkac433>
  33. Maier JA, Martinez C, Kasavajhala K *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theor Comput* 2015;11:3696–713. <https://doi.org/10.1021/acs.jctc.5b00255>
  34. Ivani I, Dans PD, Noy A *et al.* Parmbsc1: a refined force field for DNA simulations. *Nat Methods* 2016;13:55–8. <https://doi.org/10.1038/nmeth.3658>
  35. Darden T, York D, Pedersen L. Particle mesh Ewald: an *N* · log (*N*) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–92. <https://doi.org/10.1063/1.464397>
  36. Hess B, Bekker H, Berendsen HJC *et al.* LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997;18:1463–72. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H)
  37. Smith LJ, Daura X, van Gunsteren WF. Assessing equilibration and convergence in biomolecular simulations. *Proteins* 2002;48:487–96. <https://doi.org/10.1002/prot.10144>
  38. Daura X, van Gunsteren WF, Mark AE. Folding–unfolding thermodynamics of a  $\beta$ -heptapeptide from equilibrium simulations. *Proteins* 1999;34:269–80. [https://doi.org/10.1002/\(SICI\)1097-0134\(19990215\)34:3%3c269::AID-PROT1%3e3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0134(19990215)34:3%3c269::AID-PROT1%3e3.0.CO;2-3)
  39. Lavery R, Sklenar H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn* 1988;6:63–91. <https://doi.org/10.1080/07391102.1988.10506483>
  40. Zhou T, Yang L, Lu Y *et al.* DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res* 2013;41:W56–62. <https://doi.org/10.1093/nar/gkt437>
  41. Mitra R, Li J, Sagendorf JM *et al.* Geometric deep learning of protein–DNA binding specificity. *Nat Methods* 2024;21:1674–83. <https://doi.org/10.1038/s41592-024-02372-w>
  42. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2007;2:31–4. <https://doi.org/10.1038/nprot.2007.13>
  43. Petrie MV, Zhang H, Arnold EM *et al.* Dbf4 Zn-finger motif is specifically required for stimulation of Ctf19-activated origins in *Saccharomyces cerevisiae*. *Genes* 2022;13:2202. <https://doi.org/10.3390/genes13122202>
  44. Li S, Pradhan L, Ashur S *et al.* Crystal structure of FOXC2 in complex with DNA target. *ACS Omega* 2019;4:10906–14.
  45. Dai S, Li J, Zhang H *et al.* Structural basis for DNA recognition by FOXG1 and the characterization of disease-causing FOXG1 mutations. *J Mol Biol* 2020;432:6146–56. <https://doi.org/10.1016/j.jmb.2020.10.007>
  46. Tsai K-L, Sun Y-J, Huang C-Y *et al.* Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res* 2007;35:6984–94. <https://doi.org/10.1093/nar/gkm703>
  47. Li J, Dai S, Chen X *et al.* Mechanism of forkhead transcription factors binding to a novel palindromic DNA site. *Nucleic Acids Res* 2021;49:3573–83. <https://doi.org/10.1093/nar/gkab086>
  48. Pluta R, Aragón E, Prescott NA *et al.* Molecular basis for DNA recognition by the maternal pioneer transcription factor FoxH1. *Nat Commun* 2022;13:7279. <https://doi.org/10.1038/s41467-022-34925-y>
  49. Mitra R, Cohen AS, Sagendorf JM *et al.* DNAProDB: an updated database for the automated and interactive analysis of protein–DNA complexes. *Nucleic Acids Res* 2025;53:D396–402. <https://doi.org/10.1093/nar/gkac970>
  50. Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>
  51. Kundaje A, Pollard KS, Ma J *et al.* Artificial intelligence in molecular biology. *Mol Cell* 2025;85:193–8. <https://doi.org/10.1016/j.molcel.2024.12.013>
  52. Berman HM, Westbrook J, Feng Z *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235>
  53. Bandukwala HS, Wu Y, Feuerer M *et al.* Structure of a domain-swapped FOXP3 dimer on DNA and its function in regulatory T cells. *Immunity* 2011;34:479–91. <https://doi.org/10.1016/j.immuni.2011.02.017>
  54. Zhang W, Leng F, Wang X *et al.* FOXP3 recognizes microsatellites and bridges DNA through multimerization. *Nature* 2023;624:433–41. <https://doi.org/10.1038/s41586-023-06793-z>
  55. Krissinel E. Crystal contacts as nature’s docking solutions. *J Comput Chem* 2010;31:133–43. <https://doi.org/10.1002/jcc.21303>

56. Reinapae A, Ilves I, Jürgens H *et al.* Interactions between Fkh1 monomers stabilize its binding to DNA replication origins. *J Biol Chem* 2023;299:105026. <https://doi.org/10.1016/j.jbc.2023.105026>
57. Malo ME, Postnikoff SDL, Arnason TG *et al.* Mitotic degradation of yeast Fkh1 by the Anaphase Promoting Complex is required for normal longevity, genomic stability and stress resistance. *Aging* 2016;8:81030. <https://doi.org/10.18632/aging.100949>
58. Chiu TP, Li J, Jiang Y *et al.* It is in the flanks: conformational flexibility of transcription factor binding sites. *Biophys J* 2022;121:3765–7. <https://doi.org/10.1016/j.bpj.2022.09.020>