Crystal Structure of the Human FO XK1a-DNA Complex and Its Implications on the Diverse Binding Specificity of Winged Helix/Forkhead Proteins* 

Kuang-Lei Tsai§†, Cheng-Yang Huang†, Chia-Hao Chang§, Yuh-Ju Sun§, Woei-Jer Chung†, and Chwan-Deng Hsiao‡‡

From the †Institute of Molecular Biology, Academia Sinica, Taipei 115, §Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu 300, and ¶Department of Biochemistry, National Cheng Kung University, College of Medicine, Tainan 701, Taiwan

Interleukin enhancer binding factor (ILF) is a human transcription factor and a new member of the winged helix/forkhead family. ILF can bind to purine-rich regulatory motifs such as the human T-cell leukemia virus-long terminal region and the interleukin-2 promoter. Here we report the 2.4 Å crystal structure of two DNA binding domains of ILF (FO XK1a) binding to a 16-bp DNA duplex containing a promoter sequence. Electrophoretic mobility shift assay studies demonstrate that two ILF-DNA binding domain molecules cooperatively bind to DNA. In addition to the recognition helix recognizing the core sequences through the major groove, the structure shows that wing 1 interacts with the minor groove of DNA, and the H2-H3 loop region makes ionic bonds to the phosphate group, which permits the recognition of DNA. The structure also reveals that the presence of the C-terminal α-helix in place of a typical wing 2 in a member of this family alters the orientation of the C-terminal basic residues (RKKRRPR) when binding to DNA outside the core sequence. These results provide a new insight into how the DNA binding specificities of winged helix/forkhead proteins may be regulated by their less conserved regions.

The members of the extensive family of forkhead box (Fox) proteins are transcription factors important for regulating cellular proliferation, transformation, differentiation, and longevity (1–3). They are expressed in various eukaryotic organisms with multiple domains that are specific for DNA binding, trans-activation, or trans-repression (4). The Fox proteins have a unique winged helix domain composed of ~100 evolutionarily well conserved amino acids essential for DNA recognition (5). Several three-dimensional structures of this DNA binding domain (DBD), including HNF-3y (FoxA3), Genesis (Foxd3), AFX (FOXO4), FREAC11 (FOXC2), and ILF-1 (FO XK1), have been determined using x-ray crystallography or NMR spectroscopy (6–11). Most of these structures exhibit three α-helices, three β-strands, and two wing-like loops (12). The so-called winged helix DNA-binding motif of these proteins is named after the three-dimensional structure of the HNF-3y-DNA complex in which two wing-like loops (wing 1 and wing 2) and helix 3 are involved in protein-DNA interactions (13).

For many DNA-binding protein families, the differences in DNA binding specificity between family members are determined by the contact residues on recognition elements; the substitution of a DNA contact residue within family members leads to different binding specificity (14). However, transcription factors containing the winged helix-binding motif are exceptions (15). The DNA binding domains of Fox proteins exhibit a remarkable degree of sequence homology in their DNA binding region but a notable variability in their DNA recognition specificity (12). To date, more than 200 winged helix/forkhead proteins have been identified, but only two forkhead protein-DNA complexes have been reported (7, 13). Based on these two three-dimensional structures, it has been proposed that the DNA recognition helix (helix 3) is most important in DNA binding specificity. In addition, amino acid variations in the wing regions contribute to the differences in the DNA binding specificity of these proteins (12). Because of the limited number of three-dimensional structures available for these winged helix protein-DNA complexes, however, we have little concrete evidence about the diverse DNA-binding properties of these proteins (12).

Interleukin enhancer binding factors (ILFs) are transcription factors that bind purine-rich regulatory motifs in the human T-cell leukemia virus-long terminal region and the interleukin-2 (IL-2) promoter (16, 17). Previous studies suggest that ILFs bind to the regulatory sequences of the IL-2 promoter and regulate gene expression (17). DNA selection experiments identified a 6-bp motif, 5'-TAAAAC-3', as the optimal sequence for ILF binding, but this sequence varies slightly from the target gene (17). Three ILFs of 655 (ILF-1), 609 (ILF-2), and 323 (ILF-3) amino acids have been reported (16–20). There are protein sequence homologies between ILF-1 and ILF-2, including a region for potential ubiquitin-mediated degradation, a nuclear localization site, an N-glycosylation motif, and a DNA binding domain. The DNA binding domains of ILF-1 and ILF-2 are designated as FO XK1a and FO XKb, respectively (5), and are between residues 251 and 348 of the proteins. They share 35 to 89% similarity with other known members of the winged helix/forkhead family. Hence, ILF-1 and ILF-2 are classified as forkhead family members.

In a previous study (8), we determined the three-dimensional structure of the DNA binding domains of ILF-1 (ILF-DBD) by using multidimensional NMR spectroscopy and affirmed that ILF-1 is a new member of the winged helix/forkhead family. ILF-DBD, however, has a C-terminal α-helix instead of the wing 2 of typical winged helix/fork-
head proteins. This C-terminal α-helix may affect the regulation or specificity of DNA binding. In this study, we determined the three-dimensional structure of ILF-DBD complexed with a 16-bp DNA duplex and Mg\(^{2+}\) ions using x-ray crystallography. We have combined structural and biochemical evidence to provide several important insights into the DNA binding specificity of the winged helix/forkhead family proteins.

**EXPERIMENTAL PROCEDURES**

**Protein and DNA Preparation**—The coding region of 98-residue ILF-DBD was cloned into pET-21a and expressed in an *Escherichia coli* system (8). The recombinant protein contained 13 additional residues (ASMTGGQQMGRGS) at the N terminus. Selenomethionine (SeMet)-labeled recombinant protein was purified from cells grown in DBD was cloned into pET-21a and expressed in an

**TABLE 1**

| Data collection and MAD phasing statistics | FOXX1a-DNA | Se-Met-FOXX1a-DNA | Se-Met-FOXX1a-DNA |
|-------------------------------------------|------------|-------------------|-------------------|
| Crystal                                   | 1.12714    | 0.97982           | 0.98006           |
| Wavelength (Å)                            | 30-2.4     | 30-3.7            | 30-3.7            |
| Resolution (Å)                            | P6, 22     | P6, 22            | P6, 22            |
| Space group                               | a = 58.7   | a = 59.9          | a = 59.9          |
| Cell dimension (Å)                        | c = 324.9  | c = 323.7         | c = 323.7         |
| Completeness (%)                          | 99.2 (98.1) | 97.5 (92.0)       | 97.6 (91.9)       |
| Completeness (%)                          | 99.2 (98.1) | 97.5 (92.0)       | 97.6 (91.9)       |
| Completeness (%)                          | 34.3 (3.8)  | 21.7 (3.1)        | 22.8 (3.1)        |
| Completeness (%)                          | 5.3 (53.5)  | 10.7 (76.1)       | 10.6 (81.4)       |
| Refinement                                | 8.6        | 8.5               | 8.7               |
| Resolution (Å)                            | 30-2.4     | 30-3.7            | 30-3.7            |
| No. of reflections                        | 12,716     | 21.5/26.0         | 21.5/26.0         |
| No. of atoms                              | 1604       | 650               | 295               |
| Protein                                   | 39.1       | 39.1              | 39.1              |
| DNA                                       | 30.0       | 30.0              | 30.0              |
| Water                                     | 295        | 295               | 295               |
| Root mean square deviation                | 0.010      | 1.6               | 1.6               |

* a Highest resolution shell is shown in parentheses (3.8-3.7 Å for selenium and 2.5-2.4 Å for the native data set).

**Structure of FOXK1a Bound to DNA**

Crystallographic—ILF-DBD was mixed with the oligonucleotides at a 2:1 molar ratio in buffer containing 20 mM HEPES, pH 7.0, 20 mM MgCl\(_2\), and 200 mM NaCl. A protein concentration of 5 mg ml\(^{-1}\) was required for protein-DNA complex crystallization. Both native and SeMet-labeled crystals were grown at 4 °C using the hanging drop vapor diffusion method; the complex was mixed with an equal volume of reservoir solution containing 100 mM HEPES, pH 7.2, 200 mM NH\(_4\)SO\(_4\), 20% PEG3350. The crystals belonged to space group P6, 22 with cell dimensions of a = b = 58.7 Å, and c = 324.9 Å, and diffraction to 2.4 and 3.7 Å for native and SeMet-labeled crystals, respectively. The native and multiwavelength anomalous diffraction (MAD) data were collected at the Raxis-IV++ imaging plate using a synchrotron radiation x-ray source at Beamline 17B2 of the National Synchrotron Radiation Research Center in Taiwan. A single crystal was soaked in a cryoprotectant solution for 20 min before it was frozen in liquid nitrogen. The cryoprotectant solution contained the same components as the reservoir solution plus 25% glycerol. The structure of the ILF-DBD-DNA complex was determined using MAD phasing applied to the SeMet analog. The MAD data were collected at three wavelengths at 0.9798 (peak), 0.9800 (inflection), and 0.9721 Å (remote) under cryogenic conditions. All data were processed with the HKL2000 package (21).

SOLVE (22) was used to locate selenium sites and to generate the initial MAD phases at 3.7 Å. The initial phases were further improved using RESOLVE (23). XtalView (24) was used to examine electron density maps and molecular models. The native data set (2.4 Å) was used for further refinement by energy minimization and simulated annealing using CNS (25). Water molecules were added with a waterpick routine in the CNS program. The current model has an R factor of 21.5% for all reflections above 2 s between 30.0 and 2.4 Å resolution and an R\(_{\text{free}}\) of 26.0%, using 8% randomly distributed reflections. The Ramachandran plot has no violation of accepted backbone torsion angles. The helical parameters of DNA were analyzed using the CURVE (26) program. The PyMol (27) and MolMol (28) programs were used to generate the figures. The data collection and refinement statistics are shown in Table 1.

**Nucleic Acids**—Various single-stranded DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). DNA substrates were constructed by annealing complementary oligonucleotides and purified using gel electrophoresis. The sequences of the oligonucleotides, written 5’ to 3’, were as follows: S11, AAGTTGTAAC; S12, AAGTTGTAAC; S13, AAGTTGTAAC; S14, AAGTTGTAAC; S15, AAGTTGTAAC; S16, AAGTTGTAAC. All other reagents were of the highest commercially available grades.

**Electrophoretic Mobility Shift Assay**—Binding reactions were performed at 25 °C in a total volume of 10 μl in 25 mM HEPES and 1 mM MgCl\(_2\) pH 7.0. DNA substrates used in this experiment were 25 μM. After adding 5 μl of the sample loading dye containing 89 mM Tris borate, 5% glycerol, and 0.01% bromphenol blue, the resulting complexes were resolved at 4 °C on a native 6% polyacrylamide gel in TBE buffer (89 mM Tris borate, 1 mM EDTA, pH 8.3) and were visualized using 0.5 μg/ml ethidium bromide.
Structure of FOXX1a Bound to DNA

Site-directed Mutagenesis—The K3A, K45A, S50A, R52A, and K73A mutants of ILF-DBD were generated according to the QuickChange mutagenesis protocol (Stratagene, La Jolla, CA) using the pET-21a-ILF111 (98) plasmid as the template (8). These mutants were confirmed by nucleotide sequencing and mass spectra. Expression and purification of these mutant proteins were as described for wild type ILF-DBD protein.

RESULTS

Overall Structure of the ILF-DBD-DNA Complex—In the asymmetric unit of the crystal, the complex structure consisted of two FOXK1a domains (residues 625–846 of ILF-1), designated ILF-DBD1 and ILF-DBD2. These two domains bound to a 16-bp DNA duplex containing the reported core sequence, 5′-TAAACA-3′, for ILF binding (Figs. 1A and 2A). The DNA duplexes were stabilized by stacking interactions with symmetry-related DNA to form a pseudo-continuous DNA helix in the crystal. For discussion purposes, we designate residue Asp751 of ILF-1 as the Asp1 of ILF-DBD.

The structure of the complex had two ILF-DBD molecules that bound to the opposite surfaces of the major groove of the DNA helix in a head-to-tail orientation (Fig. 2B). There was no direct contact between these two protein molecules. These two ILF-DBDs interacted with DNA in a similar but not identical manner. Surprisingly, the arrangement was quite different from that of the HNF-3γ domain of HNF-3γ (8). The geometry of the DNA duplex was canonical B-DNA but with a few kinks. Briefly, the overall folding of the ILF-DBD consisted of four α-helices, three β-strands, one type I turn between the H2 and the H3 regions, and one wing (wing 1) (Fig. 2B). The architecture was similar to that of the DNA binding domain in other winged helix/forkhead proteins except for the H2-H3 turn and the C-terminal region (8). The ILF-DBD did not have the C-terminal wing extension found in the winged helix domain of HNF-3γ. The typical wing 2 structure (residues 84–91) of the canonical winged helix/forkhead proteins had been replaced by an α-helix (H4) in ILF-DBD. This α-helix lie antiparallel to and was stabilized by H1 through many hydrophobic interactions and hydrogen bonds. The residues following helix 4 (92–98) formed a coil structure. Upon complex formation, the recognition helix (H3) docked into the major groove roughly perpendicular to the DNA axis and bound extensively with the core sequence 5′-TAAACA-3′. Wing 1 and the basic residues at the C terminus of ILF-DBD interacted with the minor grooves of the 3′- and 5′-flanking regions of the core sequence, respectively. The turn in the H2-H3 loop region bound with the phosphate backbone of the DNA duplex.

The structures of the two crystallographically independent domains were superimposed with a root mean square deviation of 0.53 Å for C-α atoms of secondary structural elements and showed no distinct differences (Fig. 2C). Compared with the NMR structure of the same ILF-DBD solved in the absence of DNA, the root mean square deviation of C-α atoms was 1.57 Å (Fig. 2C). This fact revealed some slight structural variations upon DNA binding, especially in wing 1, the H2-H3 loop, and the C-terminal region of the protein. In the absence of DNA, these regions were highly disordered in solution (8).

DNA Conformation in the ILF-DBD-DNA Complex—The overall conformation of the DNA in the complex was the general B-form DNA and was bent ~19° toward ILF-DBD1. The majority of bending occurred near the major groove, which was bound by the recognition helix H3 of ILF-DBD1. In the major groove of the core sequence, the base steps 5/6, 6/7, and 9/10 were kinked in this region and had slightly higher roll angles of 8.25°, 5.7°, and 10.1°, respectively. The bending in this region may enable wing 1 of ILF-DBD1 to approach T12 and A13, both on the minor groove 3′ to the core sequence. Lys63, Ser75, and Trp77 of wing 1 interacted with the phosphate groups of the DNA and further stabilized the bending (Fig. 3). The major groove was slightly widened (~1–2 Å wider than the canonical B-DNA) at points where the DNA conformation deviated from the core sequence. The protein-phosphate interactions were localized in two phosphate backbones that formed the major groove of the core sequence. The propeller twist angle for the base pair 5′-T5-A5′ in the ILF-DBD-DNA complex was ~1°, whereas the corresponding base pair in the HNF-3γ-DNA complex was heavily propeller-twisted. Furthermore, the tilt was more negative at the TCAACC nucleotide base step in the HNF-3γ-DNA complex than that of TAAACA in the ILF-DBD-DNA complex. The helical twist per base pair varied from 26.66 to 38.83°. The DNA was 0.6% shorter than canonical B-DNA with the same number of nucleotides.
DNA Binding and Promoter Recognition at the Major Groove—The recognition helix H3 was the most highly conserved region of sequence in the forkhead family. Two ILF-DBD molecules were perpendicularly docked into the major groove of the DNA. The H3 of ILF-DBD1 centered over the TAAACA core sequence and that of ILF-DBD2 bound to the ATACA sequence (Fig. 4, A and B). Several residues within H3 bound directly and via water-mediated hydrogen bonds as well as van der Waals contacts with the bases of the major groove. Although the DNA binding arrangement was similar for both ILF-DBDs, the recognition pattern between these two molecules was slightly different (Fig. 4, A and B). In ILF-DBD1, H3 inserted into the major groove and interacted with three backbone phosphates and six bases of the DNA duplex (Figs. 3 and 4A). Asn^{49}, Ser^{50}, Arg^{52}, His^{53}, and Ser^{56} of H3 played a central role in DNA recognition. A7 was recognized by Asn^{49} through two direct hydrogen bonds. Asn^{49} also recognized T8 via a water-mediated hydrogen bond. Ser^{56} bound with base A6 using two water-mediated hydrogen bonds. Arg^{52} interacted with T10’ and T8’ through van der Waals contacts and contributed to the specificity for G9’ with a direct and a water-mediated hydrogen bond (Fig. 4A). The side chain of His^{53} protruded into the major groove and recognized bases T5 and T6’ through van der Waals force and a direct hydrogen bond (Fig. 4A). T7’ was bound by Ser^{56}, which is located at the C-terminal end of the H3, through a hydrophobic interaction. In addition to base recognition, Lys^{45}, Asn^{54}, and Ser^{56} bound with the phosphate groups of the DNA backbone to further stabilize the complex structure. The backbone phosphate groups of T8 and T11’ formed a direct hydrogen bond with Ser^{56} and Lys^{45}, respectively. However, the phosphate group of G4 interacted with Asn^{54} via a water molecule.

Ser^{50}, Arg^{52}, and His^{53} of ILF-DBD2 (Fig. 4B) recognized a specific DNA base in a manner similar to that of ILF-DBD1. Ser^{50} interacted with A11 indirectly through water-mediated hydrogen bonds. The side chain of Arg^{52} formed two direct hydrogen bonds with G14’. In addition, Arg^{52} interacted with both methyl groups of T13’ and T15’
Structure of FOXK1a Bound to DNA

FIGURE 3. Schematic diagram of protein-DNA contacts in the ILF-DBD-DNA (left) and HNF-3-γ-DNA complexes (right). Base pairs of the TAAACA sequence are in red boxes. The residues of ILF-DBD1 and ILF-DBD2 that make contact with DNA are colored in black and cyan, respectively. Hydrogen bonding and van der Waals contacts participating in DNA interactions between ILF-DBD1 and ILF-DBD2 might have been due to the DNA sequence used in this study. ILF-DBD1 showed more base pair specificity to its DNA recognition site than that of ILF-DBD2 (Fig. 3).

DNA Sequence Recognition at the Minor Groove Wing 1-DNA Interactions—In addition to the interactions with the major groove, the crystal structure showed that wing 1 of both ILF-DBDs bound to the minor groove (Fig. 2B). One of the most striking features is that wing 1 recognized the TA sequence 1 bp downstream from the TAAACA core sequence. Lys73 of wing 1 penetrated deeply into the minor groove and interacted directly with the DNA. The terminal amine of Lys73 formed direct base interaction with the N-3 atoms of A12 and A13 (Fig. 5A). Although Lys73 is highly conserved in the winged helix/forkhead family (Fig. 1), there is no binding between Lys73 and DNA in the HNF-3-γ complex. In addition, the amide group of Ser72 hydrogen-bonded with the DNA phosphate. At the stem of wing 1, the amine group of Lys63 and the amide group of Trp77 interacted with the phosphate groups of T8 and G9, respectively (Figs. 3 and 5A). The N-terminal residues (Ser67—Gln68—Glu69—Glu70) of the wing 1 were positioned away from the DNA and did not make contact with the nucleic acids.

C-terminal Basic Residues Interacted with the Minor Groove of DNA—The most interesting finding from the current study was that the basic residues (RKRRPR) following helix 4 in the C terminus of ILF-DBD appeared to be important for DNA recognition. The side chain of Arg98 in ILF-DBD1 pointed toward the DNA minor groove, and its two amine groups interacted with the O-2 atoms of nucleotides T2 and T3 through hydrogen bonds and bridging water molecules (Fig. 5, B and C). These two nucleotides are upstream from the core sequence. The side chain of Arg98 interacted with the phosphate group of G4 (Fig. 5B). In contrast, in the HNF-3-γ/DNA structure, Arg98 of HNF-3-γ formed a bidentate interaction with T1. However, Lys94 did not interact with the DNA in the ILF-DBD-DNA complex. These observations mean that ILF-1 and HNF-3-γ recognized the region upstream of the core sequence quite differently. Thus, the C-terminal region of the DNA binding domain of winged helix/forkhead proteins might contribute to DNA recognition specificity. We present this in greater detail under “Discussion.”

Mg2+-binding Site—The electron density map of the ILF-DBD/DNA structure showed a metal-binding site at the C terminus of H3 in both ILF-DBD1 and ILF-DBD2. We assumed that it was a magnesium ion because that metal was added to crystallize the complex. The Mg2+ ion was coordinated square-bipyramidally with the main chain carbonyl oxygen of Leu71, Asn74, Phe77, and Ser72, as well as with two water molecules with bond distances ranging from 2.4 to 3.2 Å. Furthermore, the Mg2+ ion also interacted with the phosphate groups of DNA through a water molecule. Similar metal ion-binding sites were found in the HNF-3-γ/DNA and IRF-2/DNA structures. They were assumed to...
FIGURE 4. Stereo views of the recognition helices and DNA interactions. A–C, recognition of the TAAACA core sequence by H3 of ILF-DBD1, ATAACA sequences by H3 of ILF-DBD2, and TCAACC sequence by H3 of HNF-3γ. Hydrogen bonding and van der Waals contacts participating in the base recognition are represented by dashed black lines. Water molecules are represented as green balls. The helix 3 is colored in red in both ILF-DBDs and colored in blue in HNF-3γ.
be a magnesium ion and a potassium ion, respectively (13, 29). The metal ion in the ILF-DBD-DNA complex may have the same function as that in the HNF-3-γ-DNA complex, where it neutralizes the helix dipole of the H3 C-terminal cap.

Mutational Analyses of ILF-DBD—Our crystal structure revealed that some residues may be important for DNA binding (Fig. 3). We previously examined the C-terminal residues (RKRRPR) of ILF-DBD crucial for DNA binding (8). Other conserved residues, such as Lys3, Ser20, Arg52, and Lys73, remain unexplored. Lys45 of ILF-DBD was not conserved in winged helix/forkhead proteins, but it interacted differently with DNA in ILF-DBD and HNF-3-γ (Fig. 4, A and C). To investigate the DNA binding activity of these residues, we constructed alanine substitution mutants and assayed them by using EMSA. As shown in Fig. 6A, the K3A, K45A, S50A, and K73A mutations led to a decrease in DNA complex formation to 40, 20, 70, and 25%, respectively. We found no band shift for the R52A mutant, which implied that no protein-DNA complex formed. We suggest that the hydrophobic force and the hydrogen bonds formed by Arg52 to the last three bases of the core sequence (TAAACA) might serve as a main step for initiating or promoting the binding of ILF-DBD to DNA.

Minimum Length of DNA Fragment Required for ILF-DBD Binding Determined by Electrophoretic Mobility Shift Assay—Our crystal structure revealed that two ILF-DBD molecules bound to a 16-bp DNA duplex (Fig. 2B). We tested oligonucleotides 11–16 bp long (S11–S16) by using EMSA to determine the minimum binding site size of the DNA fragment required for ILF-DBD binding. Mixing ILF-DBD and S14, S15, or S16 in equal mole ratios yielded significant band shifts (Fig. 6B), which implied protein-DNA complexes had been formed. At a 2:1 ILF-DBD to DNA ratio, an extra band appeared. This slower mobility band probably represents the binding of 2 ILF-DBD molecules to a DNA duplex. With a 3:1 protein to DNA ratio, this slower mobility band appeared mainly on the gel. There was, however, no band shift for the binding of ILF-DBD to S11, S12, and S13 DNA duplexes, which meant that binding was largely abolished when the size of the oligonucleotide was less than 13 bp. We had similar results when we tested other 13-bp DNA duplexes that contained only the sequence that interacts with
ILF-DBD1; this sequence was flanked by some G/Cs that acted as clamps (data not shown). These results ruled out the possibility that shorter nucleotides with high A/T content had formed unstable duplexes. We thus determined that the minimal size of oligonucleotide that permits efficient protein binding was 14 bp.

DISCUSSION

Winged helix/forkhead proteins have similar binding specificity to the core sequence, (T/C)(A/C)AA(C/T)A, and conserved amino acid sequences in the putative recognition helix (2, 3, 5, 13). This raises an intriguing question about how proteins use conserved residues to recognize distinct core sequences. Although there are more than 200 winged helix/forkhead proteins, only the complex structures of HNF-3/H9253 and Genesis-DNA (30) have been reported. HNF-3 binds to DNA specifically as a monomer. However, the length of the DNA (13-mer) used in that study was too short to show interactions between wing 1 and the DNA. A longer DNA duplex was used for the NMR solution study of the Genesis-DNA complex (30), but the structure was calculated using a straight B-form DNA template. Prior to this study, little was known about how winged helix/forkhead proteins recognize diverse DNA sequence adjacent to the core sequence. The structural analysis in this study demonstrated that winged helix/forkhead proteins recognized DNA not only with the recognition helix (H3) but also from less conserved regions as discussed below.

DNA Core Sequence Recognition by the Recognition Helix H3—The DNA binding domains of ILF, FREAC11, HNF-3γ, and Genesis recognized the core sequences TAAAACA, GTAACA, GTCAATA, and AAAATAAC, respectively (7, 9, 13, 17). Although the HNF-3γ-DNA complex structure provided the first molecular understanding about interactions between the protein and the core sequence, only Asn49 and His53 of the recognition helix interacted with G4, T5, and A7 (13) (Figs. 3 and 4). In contrast to HNF-3γ, we found important hydrogen bonds to T6 (His53), A7 (Asn49), and G9 (Arg52) and hydrophobic contacts with T5 (His53), T8 (Ser56, Arg52), and T10 (Arg52) in the ILF-DBD/DNA structure (Fig. 3). The contact patterns of the major groove by recognition helices between these two proteins were different. This result was especially surprising, because the amino acid sequences of the recognition helix in these two proteins are highly conserved. In addition, the position and orientation of the recognition helices of these two proteins were similar in the complexes. The reason for the difference was probably that the DNA sequence (GTCAACC) used in the HNF-3γ complex study differed from that of the core sequence (GTA-AAAA) in this study. Different DNA sequence may cause different DNA bending or change the network of water molecules located within the interface between the recognition helix and DNA.

In the ILF-DBD/DNA structure, we found that the base pair A6-T6 had an unusual geometry with a zero propeller. This zero propeller was not observed at the equivalent position (C6-G6) in the HNF-3γ-DNA
complex, suggesting that this geometry might have facilitated the hydrogen bonding of His\(^53\) of the ILF-DBD to T6'. In addition, DNA-binding site selection studies (17, 31–34) showed that most forkhead proteins prefer cytosine and adenine at positions 9 and 10 of the core sequence, respectively. A thymine at position 9 is tolerated by most forkhead proteins but to a lesser extent than is a cytosine. In our complex structure, Arg\(^{52}\) recognized the base pairs C9-G9' and A10-T10' through hydrogen bonding and van der Waals contact, respectively. However, this important recognition did not occur in the HNF-3-γ-DNA complex. Comparing the 5'-T10-pG9'-3' step of the ILF-DBD/DNA structure with that of the 5'-T "pG"-3' step of the B-type DNA showed that T10' in the DNA-bound structure was displaced into the major groove and approached the side chain of Arg\(^{52}\). Additionally, our structure showed that the DNA was kinked toward ILF-DBD1 with a high roll angle of \(-10°\) at this step. These observations suggested that the recognition of T10' by Arg\(^{52}\) was significant. Substituting alanine for Arg\(^{52}\) dramatically reduced ILF-DBD binding to the DNA (Fig. 6A), which was consistent with another positional equivalent mutation of R127H in FOXC1 that led to a significant disruption in the DNA binding affinity of the protein (35). Thus, we hypothesize that Arg\(^{52}\) is essential for DNA binding and is important for core sequence recognition.

**H2-H3 Loop Region May Regulate the ILF-DNA Interaction**—The DNA binding specificity of winged helix/forkhead proteins is influenced by residues located in the H2-H3 loop region (9, 32, 34). These residues have variable sequences. The H2-H3 region is formed by a short helix, a random-coil segment, and a 3\(\alpha\) helix in Genesis, ILF-DBD/DNA, and HNF-3-γ/DNA, respectively. It has been proposed (15) that the these structures may regulate the relative presentation of helix 3, thereby leading to different binding specificities.

Although the ILF-DBD and HNF-3-γ share a high degree of sequence identity within the DNA recognition helix, H3, these two proteins have slightly different DNA core sequence specificities. The core sequences are 5'-TAAACA-3' for the ILF-DBD and 5'-GTAAACA-3' for HNF-3-γ. The H2-H3 region of these two proteins has a segment of highly conserved residues (TYR\(^{27}\)/PHR\(^{27}\)-PRO-Tyr-Tyr-Arg\(^{44}\)). However, the residues before and after this segment are very diverse (Fig. 1). Even though the orientation of H3 in these two proteins is not different, we detected obvious structural differences in the H2-H3 loop region, namely residues that are closed to H3 in particular. Thus, it is possible that residues 42–46 might have an important effect on the DNA binding specificity of winged helix/forkhead proteins.

In the ILF-DNA complex, the side chain of Lys\(^{45}\) from ILF-DBD1 protrudes upward to form a salt bridge with the phosphate group of T11' (Fig. 4A), and the side chain of Lys\(^{56}\) from ILF-DBD2 forms a direct contact with the base T15' (Fig. 4B). These interactions further bend the DNA toward the recognition helices (H3) and form specific base recognition. Arg\(^{52}\) of DBD2 and Arg\(^{46}\) of HNF-3-γ protrude in an opposite direction and forms an ionic interaction with the phosphate group of C6 (Fig. 4C). It is consistent with the reported structures of the DNA-protein complexes that the neutralization of charge by lysines or arginines will influence the bending of DNA (36, 37).

It is noteworthy that the orientations of Arg\(^{52}\) in ILF-DBD and Arg\(^{46}\) in HNF-3-γ may cause a difference in the electrostatic potentials on the protein surfaces. Consequently, the electrostatic interaction of these two proteins with the DNA phosphate backbone may be affected, and the presentation of H3 to the major groove of the DNA may be modulated. Furthermore, in study of FREAC1-7 (32), the specificity for 5'-GTAAATA-3' versus 5'-GTAAACA-3' is determined by a short stretch of amino acid residues at the junction between the H2-H3 turn and the first three residues of H3. Therefore, we hypothesize that these residues cause the slight variations in the core sequence binding properties between the ILF-DBD and HNF-3-γ.

**The Role of Wing 1 in the Winged Helix/Forkhead Proteins**—Because a short DNA sequence (13 bp) was used in the HNF-3-γ complex (13), the interaction between wing 1 and DNA was not observed. A structural comparison between the DNA-bound and free forms of ILF-DBD reveals that wing 1 undergoes a major structural rearrangement upon DNA binding. Although wing 1 is mostly disordered in the absence of DNA (8), it is stabilized by protein binding to the minor groove of the DNA through direct hydrogen bonds. In the ILF-DBD-DNA complex structure, Lys\(^{73}\) of wing 1 interacts with the bases at the minor groove of the 3'-flanking core sequence. The K73A mutant significantly loses the DNA binding ability. These results indicate that wing 1 of the ILF-DBD is important for DNA binding and may play a role in DNA recognition. In addition, wing 1 and DNA interactions may stabilize a DNA bending toward the protein and help H3 recognize the core sequence. Therefore, the ILF-DBD-DNA complex structure may provide a model to show how winged helix/forkhead proteins use wing 1 to bind with DNA in the minor groove.

Sequence alignment shows that ILF has a negatively charged Glu\(^{70}\) in wing 1 (Fig. 1). Interestingly, the corresponding residue is a positively charged lysine residue in HNF-3-γ or Freac11, and a polar asparagine residue in Genesis. This difference may also contribute to DNA binding in other wing helix/forkhead proteins. However, this hypothesis needs to be confirmed by further biochemical studies.

**Implications of DNA Binding Specificity Recognized by C-terminal Basic Residues**—A sequence comparison of the DNA binding domain of winged helix/forkhead proteins shows that the C terminus is one of the most diverse regions (Fig. 1). DNA binding specificities of winged helix/forkhead proteins are strongly influenced by the C-terminal part of the DNA binding domain (32). Instead of the typical wing structure of the canonical winged helix/forkhead proteins, the corresponding region (residue 84–91) in ILF had an α-helical structure.

Protein truncation and gel retardation studies (8) showed that deletion of the C-terminal six residues (residues 93–98) seriously reduced the DNA binding ability, which suggests that this region is important for protein-DNA interaction. In the present study, we found that helix 4 of the ILF-DBD altered the orientation of the C-terminal basic residues (RKRPRP\(^{96}\)) and made contacts to the minor groove that were different from those in the HNF-3-γ-DNA complex (Fig. 5C). This indicated that the C-terminal region of the DNA binding domain in the winged helix/forkhead family mediated the DNA binding specificities of 5’-flanking core sequences.

**Cooperative Binding at the Two DNA Sites**—This study provides the first view of cooperative binding of the ILF-DBD to DNA that contains the core sequence. The cooperativity of the ILF-DBD can arise through DNA conformability in the absence of strong protein-protein interactions. The structural evidence presented here showed that there were two ILF-DBDs per DNA duplex, whereas the biochemical data suggested cooperative binding. EMSA experiments showed that DNA at least 14 bp long was required for the two ILF-DBD molecules to bind optimally (Fig. 6B). It is of particular interest that when the second ILF-DBD molecule bound to the ILF-DBD-DNA complex, it induced DNA distortion that narrowed the major groove by bending the DNA helix 19° toward the protein. This DNA structure deformation by the second ILF-DBD binding seemed essential for stabilizing the ILF-DBD-DNA complex and might therefore explain why the ability of the ILF-
DBD to bind to the 11-, 12-, and 13-bp-long DNA sequences was severely impaired (Fig. 6B).

The DNA binding domains of transcription factors are frequently sufficient to mediate cooperative binding at composite regulatory sites. It was shown previously that the DNA core recognition sequence is essential but not sufficient for the binding of winged helix/forkhead proteins (10). In the case of ILF-DBD, the structural and functional analysis of the complex revealed that the second ILF-DBD bound to the ILF-DBD-DNA complex was likely a key step for forming a stable complex. The unique DNA binding mode found in ILF may relate to the fact that one ILF-DBD molecule binds to DNA with low affinity, and two ILF-DBD molecules bind to DNA with high affinity. Consequently, the second ILF-DBD molecule bound to DNA may serve as a regulatory element for transcription of the human T-cell leukemia virus-long terminal region or interleukin-2. In this case, the release of the second ILF-DBD molecule from the complex would destabilize the binding of the first ILF-DBD molecule and dissociate it from the core sequence. The transcription would therefore stop. This hypothesis has yet to be shown in vivo.

Insights into the Diverse Binding Specificities of Winged Helix Proteins—The winged helix/forkhead proteins shared a common fold with diverse DNA binding modes (12). In most cases, use of the recognition helix to recognize bases in the major groove is conserved. The extra structural elements, such as the wings and the H2-H3 loop, provide additional contacts with the DNA backbone. In the case of heat shock transcription factor (HSF), its wing 1 mediates dimerization rather than contacting the DNA (38). Although the nonconserved elements make less contact with DNA, they are important for regulating how the recognition helix makes specific base contacts with DNA. Domain swapping and mutagenesis studies on winged helix proteins revealed that the binding specificity of winged helix proteins depends on the H2-H3 loop (6, 15, 34). The structure of the FOXK1a-DNA complex provides the first evidence that the H2-H3 loop is in contact with DNA and that the wing 1 and the C-terminal tail also make specific base contacts. This study provides a new insight into how the nonconserved regions of winged helix/forkhead proteins may regulate their DNA binding specificities.

Acknowledgments—We thank Dr. Ming F. Tam for discussion and critical reading of the manuscript. We gratefully acknowledge access to the synchrotron radiation beamline 17B2 at the National Synchrotron Radiation Research Center in Taiwan.

REFERENCES
1. Hromas, R., and Costa, R. (1995) Crit. Rev. Oncol. Hematol. 20, 129–140
2. Kaufmann, E., and Knochel, W. (1996) Mech. Dev. 57, 3–20
3. Lai, E., Clark, K. L., Burley, S. K., and Darnell, J. E., Jr. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 10421–10423
4. Granadino, B., Perez-Sanchez, C., and Rey-Campos, J. (2000) Curr. Genomics 1, 353–382
5. Kaestner, K. H., Knochel, W., and Martinez, D. E. (2000) Genes Dev. 14, 142–146
6. Marsden, I., Jin, C., and Liao, X. (1998) J. Mol. Biol. 278, 293–299
7. Jin, C., Marsden, I., Chen, X., and Liao, X. (1999) J. Mol. Biol. 289, 683–690
8. Liu, P. P., Chen, Y. C., Li, C., Hsieh, Y. H., Chen, S. W., Chen, S. H., Jeng, W. Y., and Chuang, W. J. (2002) Proteins 49, 543–553
9. van Dongen, M. J., Cederberg, A., Carlson, P., Enerback, S., and Wikstrom, M. (2000) J. Mol. Biol. 296, 351–359
10. Weigelt, J., Climenti, I., Dahlman-Wright, K., and Wikstrom, M. (2001) Biochemistry 40, 5861–5869
11. Sheng, W., Rance, M., and Liao, X. (2002) Biochemistry 41, 3286–3293
12. Gajiwala, K. S., and Burley, S. K. (2000) Curr. Opin. Struct. Biol. 10, 110–116
13. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) Nature 364, 412–420
14. Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000) Genome Biol. http://genomebiology.com/2000/1/1/reviews/001
15. Marsden, I., Chen, Y., Jin, C., and Liao, X. (1997) Biochemistry 36, 13248–13255
16. Li, C., Lai, C. F., Sigman, D. S., and Gaynor, R. B. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 7739–7743
17. Nirula, A., Moore, D. I., and Gaynor, R. B. (1997) J. Biol. Chem. 272, 7736–7745
18. Li, C., Luscombe, N. M., Austin, S. E., and Thornton, J. M. (1999) J. Biol. Chem. 274, 40528–40541
19. Buass, F. W., Lee, K., Edelhoff, S., Distech, C., and Braun, R. E. (1999) Mamm. Genome 10, 451–456
20. Tang, J., Luo, P. N., and Herschman, H. R. (2000) J. Biol. Chem. 275, 19866–19876
21. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
22. Terwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D. Biol. Crystallogr. 55, 849–861
23. Terwilliger, T. C. (2003) Acta Crystallogr. Sect. D. Biol. Crystallogr. 59, 38–44
24. McRee, D. E. (1999) J. Struct. Biol. 125, 156–165
25. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D. Biol. Crystallogr. 54, 905–921
26. Lavery, R., and Sklenar, H. (1989) J. Biomol. Struct. Dyn. 6, 655–667
27. DeLano, W. L. (2002) The PyMOL User’s Manual. Delano Scientific. San Carlos, CA
28. Koradi, R., Billeter, M., and Wuthrich, K. (1996) J. Mol. Graphics 14, 51–55
29. Fujii, Y., Shimizu, T., Kusumoto, M., Kyogoku, Y., Taniguchi, T., and Hoshikawa, T. (1999) EMBO J. 18, 5028–5041
30. Jin, C., and Liao, X. (1999) J. Mol. Biol. 292, 641–651
31. Yang, Q., Bassel-Duby, R., and Williams, R. S. (1997) Mol. Cell. Biol. 17, 5236–5243
32. Pierreau, S., Hellqvist, M., Samuelsson, L., Enerback, S., and Carlsson, P. (1994) EMBO J. 13, 5002–5012
33. Roux, J., Pictet, R., and Grange, T. (1995) DNA Cell Biol. 14, 385–396
34. Overdier, D. G., Porcella, A., and Costa, R. H. (1994) Mol. Cell. Biol. 14, 2755–2766
35. Saleem, R. A., Banerjee-Basu, S., Berry, F. B., Baxevanis, A. D., and Walter, M. A. (2003) Hum. Mol. Genet. 12, 2993–3005
36. Strauss, J. J., and Maller, J. L. (1994) Science 266, 1829–1834
37. Strauss-Soukup, J. K., Rodrigues, P. D., and Mahar, J. L. (1998) Biophys. Chem. 72, 297–306
38. Littlefield, O., and Nelson, H. C. (1999) Nat. Struct. Biol. 6, 464–470