Structure *14*, 159–166, January 2006 DOI 10.1016/j.str.2005.10.005

Structure of the Forkhead Domain of FOXP2 Bound to DNA

James C. Stroud,[1,4,5] Yongqing Wu,[1,4,5] Darren L. Bates,[1,4] Aidong Han,[1,4] Katja Nowick,[2] Svante Paabo,[2] Harry Tong,[3,*] and Lin Chen[1,4,*]

[1] Department of Chemistry and Biochemistry
University of Colorado at Boulder
Boulder, Colorado 80309

[2] Max Planck Institute for Evolutionary Anthropology
D-04103 Leipzig
Germany

[3] Australian Synchrotron Research Program/
Consortium for Advanced Radiation Sources
Argonne National Laboratory
Building 434-B
9700 South Cass Avenue
Argonne, Illinois 60439

Summary

FOXP (FOXP1–4) is a newly defined subfamily of the forkhead box (FOX) transcription factors. A mutation in the FOXP2 forkhead domain cosegregates with a severe speech disorder, whereas several mutations in the FOXP3 forkhead domain are linked to the IPEX syndrome in human and a similar autoimmune phenotype in mice. Here we report a 1.9 Å crystal structure of the forkhead domain of human FOXP2 bound to DNA. This structure allows us to revise the previously proposed DNA recognition mechanism and provide a unifying model of DNA binding for the FOX family of proteins. Our studies also reveal that the FOXP2 forkhead domain can form a domain-swapped dimer, made possible by a strategic substitution of a highly conserved proline in conventional FOX proteins with alanine in the P subfamily. Disease-causing mutations in FOXP2 and FOXP3 map either to the DNA binding surface or the domain-swapping dimer interface, functionally corroborating the crystal structure.

Introduction

Forkhead box (FOX)-containing transcription factors are unified by sequence similarity within an approximately 90 amino acid winged-helix DNA binding domain from which the FOX family derives its name (Mazet et al., 2003). The diverse roles of FOX (human protein names are used throughout) family members in development are underscored by the fact that mutations in several members of the family are linked to congenital defects, including familial glaucoma and Axenfeld-Rieger anomalies (FOXC1) (Lehmann et al., 2000; Mears et al., 1998; Mirzayans et al., 2000; Nishimura et al., 1998, 2001), lymphedema (FOXC2) (Fang et al., 2000; Finegold et al., 2001), T cell immunodeficiency (FOXN1) (Frank et al., 1999), and thyroid agenesis with cleft palate and choanal atresia (FOXE1) (Clifton-Bligh et al., 1998). Often, these mutations are within the well-conserved forkhead domain (Carlsson and Mahlapuu, 2002), demonstrating the importance of DNA recognition and binding to the function of FOX proteins.

FOXP (FOXP1–4) is a newly defined subfamily of the FOX transcription factors that contains several recognizable sequence motifs, including a glutamine-rich region, a zinc finger, a leucine zipper, and a highly divergent forkhead domain (Lai et al., 2001; Li and Tucker, 1993; Shu et al., 2001). As seen in several other FOX proteins linked to human developmental disorders (Lehmann et al., 2003), the majority of disease-causing mutations in FOXP2 and FOXP3 occur in the forkhead domain. For instance, an arginine-to-histidine missense mutation (R553H) in the FOXP2 forkhead domain has been linked to a severe speech and language disorder (Lai et al., 2001). A deletion of the forkhead domain arising from a frame-shift mutation in the FOXP3 protein in mouse is linked to the autoimmune disorder scurfy (Brunkow et al., 2001; Schubert et al., 2001). A similar congenital disease in human is known as IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome) (Bennett et al., 2001; Wildin et al., 2001). Afflicted individuals display a variety of symptoms that include anemia, insulin-dependent diabetes, chronic diarrhea, and dermatitis (Levy-Lahad and Wildin, 2001). The similarities between human IPEX and mouse scurfy phenotypes are reflected by the fact that several mutations in the forkhead domain of the human FOXP3 gene have been linked to IPEX (Bennett et al., 2001; Wildin et al., 2001).

To address the mechanisms by which these disease-related mutations disturb FOXP function, we have determined the structure of the human FOXP2 forkhead domain bound to DNA containing a FOXP binding site (Schubert et al., 2001; Wang et al., 2003) (Table 1). Our results show that disease-causing mutations in the FOXP family map to the DNA binding interface and to a dimer interface formed by domain swapping. Domain swapping can be disrupted by replacing an alanine conserved in the FOXP family with a proline that is highly conserved in other FOX families. These results suggest that domain swapping is a unique structural feature of the FOXP forkhead domain and thus may be functionally relevant. Additionally, the high resolution of the data allows us to reinterpret earlier, lower resolution studies and propose a general model of DNA recognition by forkhead-containing proteins.

Results and Discussion

Overall Structure

The asymmetric unit (ASU) contains six copies of the FOXP2 forkhead domain and two double-stranded segments of DNA (Figure 1A). Although all six copies of FOXP2 are identical in sequence, two copies exist in a monomeric form (Figure 1A, labeled 1 and 2) and the four other copies exhibit domain swapping (described in

*Correspondence: lin.chen@colorado.edu (L.C.); tong@cars.uchicago.edu (H.T.)
[4] Lab address: http://keres.colorado.edu/
[5] These authors contributed equally to this work.

Table 1. Statistics of Crystallographic Analysis

Data Collection	
Resolution (Å)	40.0–1.90
R_{sym} (%)[a]	0.055
Completeness (%)[b]	99.7 (99.1)
I/σ[b]	28.46 (4.8)
Refinement	
Resolution (Å)	40.0–1.90
R factor[b,c]	0.217 (0.243)
R_{free}[b,c]	0.235 (0.278)
Rms deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.1
Average B factor (Å^2)	37.6

[a] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity and <I> is the statistically weighted average intensity of multiple observations of symmetry-related reflections.
[b] Numbers in parentheses are for the outer shell.
[c] R factor = $\Sigma ||F_o| - |F_c||/\Sigma |F_o|$, where $|F_o|$ and $|F_c|$ are observed and calculated structure factor amplitudes, respectively. R_{free} is calculated for a randomly chosen 9.1% of reflections.

detail below; Figure 1A, labeled 3–6). The two FOXP2 monomers bind intimately to equivalent sites on the two segments of DNA (described below), whereas the two swapped dimers loosely associate with DNA. The DNA-bound monomeric form folds into the canonical winged-helix motif characteristic of the FOX family (Clark et al., 1993) (Figure 1B). Its core is comprised of three stacking α helices (H1, H2, and H3) capped at one end by a three-stranded antiparallel β sheet (S1, S2, and S3). The turn between H2 and H3 contains a 3_{10} helix (H4) as seen in other FOX proteins (Clark et al., 1993; Jin et al., 1999; Liu et al., 2002; Weigelt et al., 2001).

Between strands S2 and S3, conventional FOX proteins contain a 5–7 amino acid insert, called wing 1. However, in FOXP2 this insert is truncated, resulting in a simple type I turn that joins strands S2 and S3 (Figure 1C). The C-terminal region also distinguishes the FOXP subfamily from most other FOX proteins. In FOXA3 this region forms an extended loop, called wing 2 (W2), that contacts DNA extensively (Clark et al., 1993). The corresponding region in FOXP2 forms a helix (H5) that runs atop H1 and terminates at the DNA phosphate backbone (Figure 1B). A similar helix H5 is also observed in the NMR structures of FOXD1 and FOXK1a (ILF-1), but the sequences and trajectories of these helices are notably different from that of FOXP2 (Jin et al., 1999; Liu et al., 2002). The heightened variability of the W1 and W2 regions relative to the rest of the forkhead domain across all FOX subfamilies suggests that the wings may have specialized functions within each subfamily.

DNA Recognition

DNA recognition by FOXP2 is mediated predominantly by helix H3 (Figure 2). Asn550 forms bidentate hydrogen bonds with Ade10, whereas His554 and Arg553 form direct or water-mediated hydrogen bonds with Thy10′ and Thy11′, respectively. The main chain and side chain atoms of Arg553, His554, Ser557, and Leu558 also make extensive van der Waals contacts to Cyt8, Gua8′, Thy9′, Thy10′, Thy11′, Ade12′, and Ade13′. A number of aromatic or hydrophobic residues from helix H1 (Tyr509), H2 (Leu527 and Tyr531), and strand S3 (Trp573) interact extensively with both helix H3 and the sugar-phosphate backbone, thereby wedging helix H3 deep into the major groove and stabilizing the protein-DNA complex (Figure 2A). In the periphery of the FOXP2/DNA interface, residues from the N and C termini (Arg504, Thr508, Arg583, and Arg584), including the main chain amide of Tyr509 at the N-terminal end of helix H1 and residues from S2 (e.g., Arg564), make hydrogen bonds, van der Waals contacts, and electrostatic interactions with the DNA backbone, providing further stability to the FOXP2/DNA complex.

Based on the major groove contacts, the DNA binding site of FOXP2 can be defined as 5′-C**AAAT**T-3′ (the core binding sequence is in bold) (Figure 2B), which is similar to that derived from in vitro selection (5′-A[C/T]**AAAT**A-3′) (Wang et al., 2003). DNA binding by FOXP2 shares a similar global structure with the FOXA3/DNA complex (Clark et al., 1993). Surprisingly, the binding site determined for FOXA3 in the previous crystallographic study was 5′-TAAGTCAACC-3′ (underlined), significantly different from that seen here for FOXP2. The DNA binding mechanism of FOXA3 derived from the early study was also significantly different from that described for FOXP2 above. However, upon careful examination, we found that conserved residues on helix H3 of FOXA3 (corresponding to Arg553, Asn550, His554, and Ser557 in FOXP2) bind a DNA region (bold) in 5′-TAAGT**CA AC**C-3′ similarly to their counterparts in FOXP2. Notably, these residues are highly conserved among all known members of the FOX family (Figure 1C). We propose that this region is a cryptic FOX binding site in the FOXA3/DNA complex (Clark et al., 1993). The revised interpretation of the DNA binding mechanism is not a result of the different DNA sequences used in the FOXA3/DNA complex and the FOXP2/DNA complex, as conserved DNA binding residues of FOXA3 and FOXP2 engage in similar DNA binding interactions in the two complexes. Thus, based on the common features of protein/DNA interactions in the FOXA3/DNA complex and the present structure, we are able to redefine the FOX binding sequence (5′-C**AAAT**T-3′) at the structural level, which is consistent with the footprinting of FOXA3 and biochemical data on the binding site of a number of FOX proteins, including FOXK1 (5′-T**AAAC**A-3′), FOXC2 (5′-GT**AAAC**A-3′), and FOXD1 (5′-A**AAAT**AAC-3′) (Costa et al., 1989; Jin et al., 1999; Liu et al., 2002; Nirula et al., 1997; van Dongen et al., 2000). A major difference in DNA binding between FOXP2 and FOXA3 is at the peripheral protein/DNA interface, where FOXA3 uses two loops (W1 and W2) to bind the DNA backbone and minor groove extensively. The corresponding loops in FOXP2 are much shorter and make limited DNA contacts (Clark et al., 1993). Consistent with these structural observations, the forkhead domain of FOXP2 binds DNA with a lower affinity than that of FOXA3 (Clark et al., 1993; Li et al., 2004).

Compared with most sequence-specific transcription factors, an unusual feature of DNA binding by FOXP2 is its extensive utilization of van der Waals contacts and a relatively small number of hydrogen bonds to bases in the major groove. This shape recognition may allow FOXP2 to bind a broad range of sequences in different promoter contexts as long as the DNA maintains the few hydrogen bond determinants in the core region of

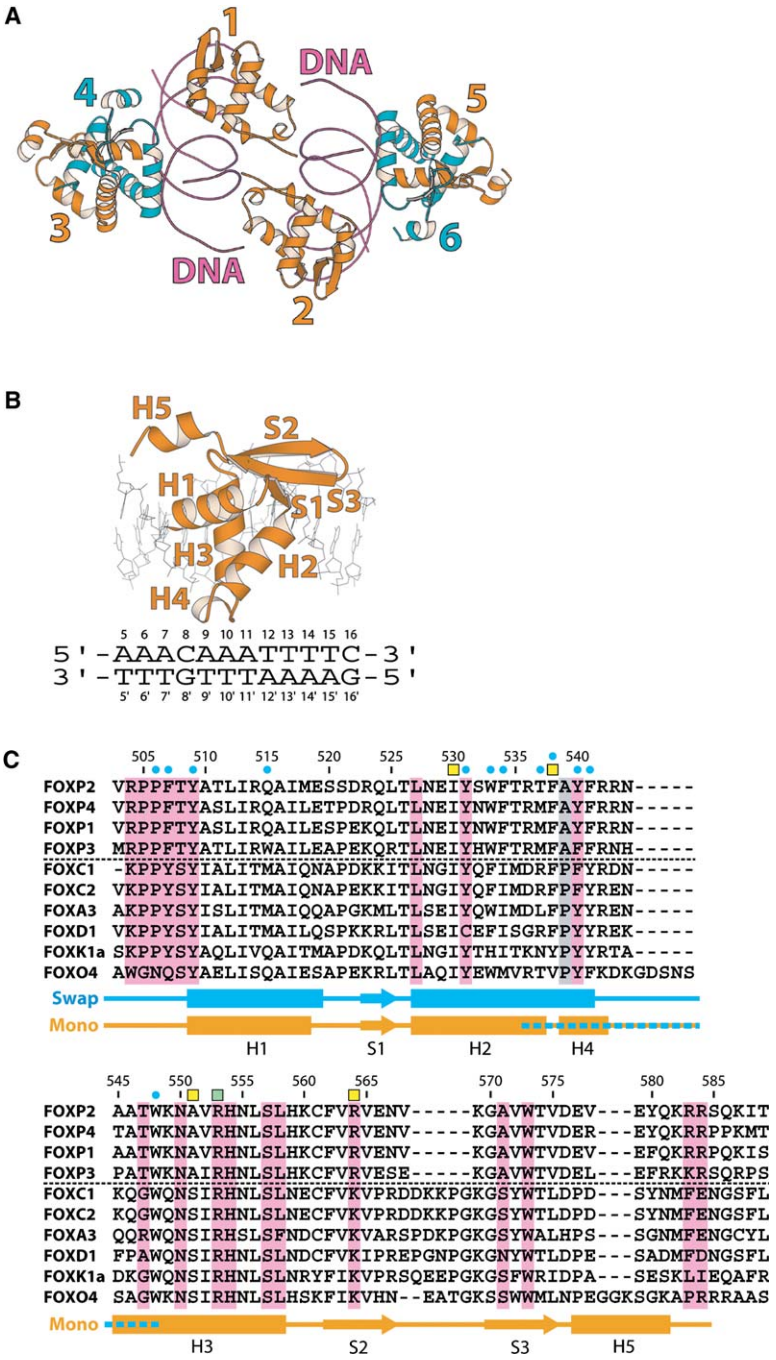

Figure 1. Overall Structure

(A) The asymmetric unit. FOXP2 molecules are shown as ribbon drawings. Molecules within the swapped dimers are both orange (3 and 5) and cyan (4 and 6). Monomers (1 and 2) are orange. The DNA phosphate backbone is shown as a coil in magenta.

(B) Ribbon drawing of FOXP2 in the monomeric form bound to DNA. The sequence of the region of DNA pictured is below the complex. DNA is shown as wire frame.

(C) Sequence alignment. FOXP proteins are separated from other FOX proteins by a dashed line. Differences in secondary structure in the first half of the protein (residues 503–544) between the monomer (Mono, orange) and the dimer (Swap, cyan) are shown below the sequence. The second half (residues 545–584) has the same secondary structure in the monomer (shown in orange) and dimer (not shown). Residues that require significant backbone changes between the monomeric and swapped forms are indicated as a cyan dash superimposed on the monomeric secondary structure representation. Residues involved in DNA binding (shaded in magenta) and intermolecular interactions in the swapped dimer (cyan circles above the sequence) are highlighted. The arginine (R553) linked to speech disorder is indicated by a green filled box. Residues homologous to those of FOXP3 linked to autoimmune diseases are indicated by yellow filled boxes. The alanine (A539) found to be critical for swapping is shown with shaded background.

the binding site and has shape complementary to the DNA binding surface of FOXP2. Because DNA-contacting residues on H3 are almost absolutely conserved (Figure 1C), this DNA binding behavior is likely common to all FOX proteins, which do not recognize a single consensus sequence but rather a degenerate pattern: 5′-RYMAAYA-3′ (R = A or G; Y = C or T; M = A or C) (Carlsson and Mahlapuu, 2002). Although there is evidence that a leucine zipper motif preceding the forkhead domain may be required for high-affinity DNA binding by FOXP proteins (Li et al., 2004), our studies here suggest that the isolated FOXP forkhead domain is capable of specific DNA binding based on a number of observations. First, the FOXP2 forkhead domain binds its cognate site in two independent complexes of the crystal asymmetric unit. Second, the detailed binding interactions observed at the FOXP2/DNA interface are conserved in the FOXA3/DNA complex. Finally, the DNA binding mechanism derived from the FOXP2/DNA complex is consistent with biochemical data (see above). However, given the short recognition sequence and relatively weak DNA binding affinity of the forkhead domain of FOXP (Li et al., 2004), it is likely that specific DNA binding by FOXP proteins in vivo will be facilitated by protein/protein interactions in higher order transcription factor complexes (Bettelli et al., 2005; Li et al., 2004).

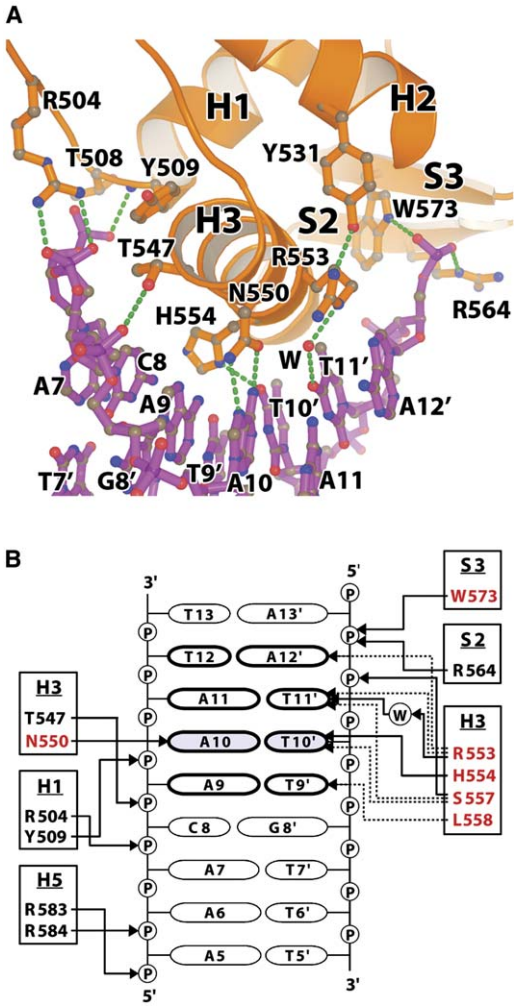

Figure 2. DNA Recognition

(A) Detailed interactions between the forkhead domain of human FOXP2 (orange) and its cognate DNA site (magenta). The DNA and protein residues are drawn as a stick model.

(B) Schematic of interactions between FOXP2 and DNA. DNA is represented as a ladder with bases as ovals and labeled according to the text (the core sequence is highlighted by thick lines). The backbone phosphates are represented as circles with the letter P inside. Hydrogen bonding interactions are solid arrows while van der Waals interactions are dashed arrows. Secondary structure elements of FOXP2 are boxed and labeled. Highly conserved residues that contribute to DNA specificity in the FOXP2/DNA and FOXA3/DNA complexes are highlighted in red. A water molecule is represented as a circle with a W inside.

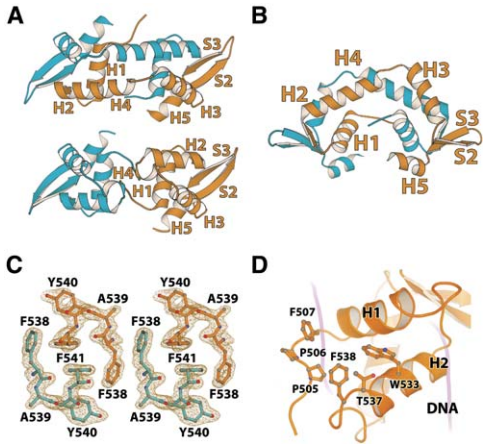

Figure 3. Structure of Domain Swapping

(A) Top view of domain swapping in FOXP2. The two FOXP2 forkhead domains are represented as orange (labeled) and cyan. Bottom panel: two monomers have been placed in positions to illustrate the rearrangements required for swapping.

(B) Side view of domain swapping. This view is rotated 90° around the horizontal axis relative to (A).

(C) Stereodiagram of electron density around the core of the swapped dimer interface.

(D) A number of hydrophobic residues exposed on the surface of the monomeric FOXP2 bound to DNA. These residues include Pro506, Phe507, Phe538, and Trp533. These residues become buried in the domain-swapped dimer.

Domain Swapping

A striking structural feature of FOXP2 is its propensity to form a domain-swapped dimer wherein two monomers of FOXP2 exchange helix H3, and strands S2 and S3 (Figure 3A). The swapping buries an additional 804 Å^2 of solvent-accessible surface and creates a semicircular arch with a pseudo 2-fold symmetry (Figure 3B). Helices H2, H4, and H3 form the convex surface of the arch, while helices H1 and H5 form the concave surface. An elaborate interaction network of aromatic residues spans the core of the swapped dimer. These residues include Phe507, Tyr509, Tyr531, Trp533, Phe534, Phe538, Tyr540, Phe541, and Trp548 from both dimers. Figure 3C shows part of this interaction network around the 2-fold axis of the swap. Here, Phe541 stacks face to face with its pseudosymmetry mate, Phe538 packs face to edge against Phe541 of its dimer partner, and Tyr540 stacks edge to face against Phe541 of the same FOXP2 copy. The swapped dimer buries several hydrophobic residues that are exposed in the DNA-bound monomeric species, including Pro506, Phe507, Trp533, and Phe538. However, these residues are highly conserved in monomeric FOX proteins (Clark et al., 1993) (Figure 3D). Thus, burial of exposed hydrophobic residues in FOXP2 must be supplemented by other factors that contribute to its propensity for domain swapping.

Domain swapping in FOXP2 is a result of the extension of helix H2 through the turn connecting H2 to H3 (Figures 3A and 4A), which creates a single straight 15 amino acid α helix in place of the shorter helices of H2 and H4. This region, corresponding to residues 538–541 (FAYF) in FOXP2, is highly conserved in all FOX proteins, except for residue Ala539. In classical FOX proteins, this position is occupied by a proline (Figure 1C), which most likely prevents the merging of helices H2 and H4 and therefore precludes domain swapping. This proline is strategically replaced by an alanine residue in all FOXP members, suggesting that domain swapping is a common feature in the FOXP family. Thus, contrary to many cases of 3D domain swapping observed under nonphysiological conditions or with artificially mutated proteins, it seems that domain swapping is an adaptive structural feature of the P branch of FOX proteins. Consistent with this hypothesis, we have shown that the forkhead domain of FOXP2 exists as both a monomer and dimer in solution with a slow exchange rate (Figure 4B) (see Experimental Procedures for further details).

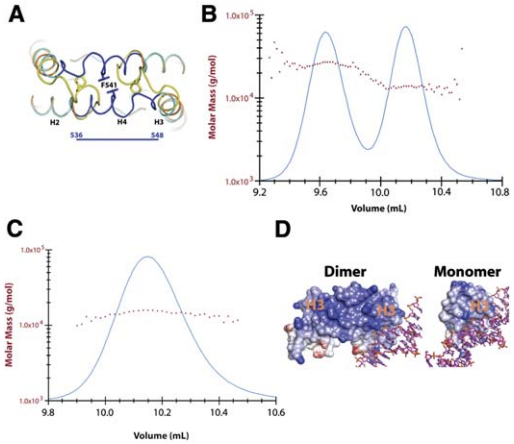

Figure 4. Biochemical Analysis and Functional Implication of Domain Swapping

(A) Superposition of the domain-swapped FOXP2 dimer (cyan) and the monomer (yellow) showing the region (blue) that undergoes significant backbone conformational changes. This region corresponds to residues 536–548 in human FOXP2, indicated by a dashed line in Figure 1C.

(B) Multiangle light scattering (MALS) analysis of the wild-type human FOXP2 (residues 503–584). The profile shows two discrete peaks corresponding to monomer (13.8 KD) and dimer (26.2 KD). Blue lines: refractive index signal profile: MALS measurement of mass at that point of the elution. The center of the peak gives the greatest signal-to-noise for the measurement of mass.

(C) MALS analysis of the Ala539Pro mutant of human FOXP2 (residues 503–584) showing a single monodispersive peak corresponding to monomer (14.1 KD).

(D) Electrostatic surface potential of the domain-swapped FOXP2 dimer (left). The DNA binding helix (H3) of one monomer in the domain-swapped dimer inserts into the DNA major groove in a similar manner to that seen in the monomer/DNA complex (shown on the right for comparison), although its interaction with DNA is loose due to the noncognate DNA sequence (not shown). The other monomer can presumably bind a separate DNA substrate (Figure 5A) and the relatively positive surface potential (blue) on top of the arch may facilitate the binding of two strands of DNA (Figure 5B).

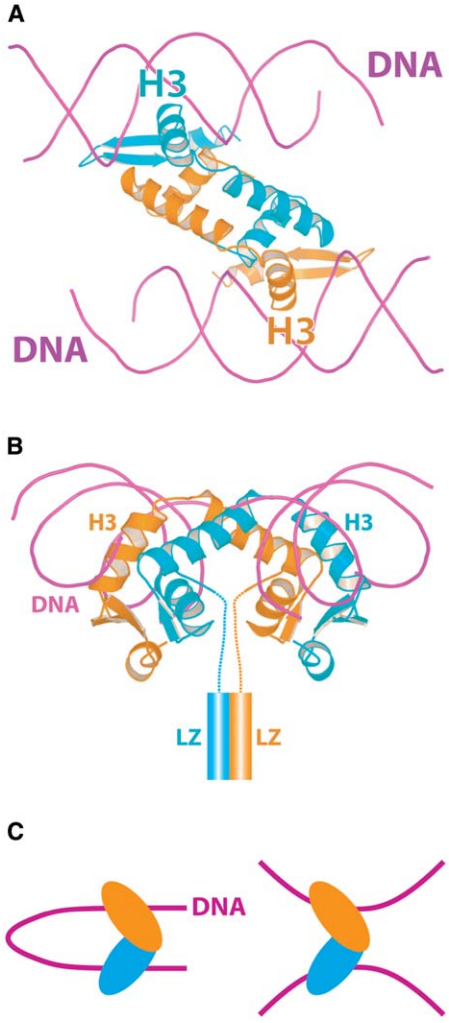

Figure 5. A Specialized Function of FOXP Proteins May Be to Promote the Assembly of Higher Order Protein/DNA Complexes

(A) A model of the domain-swapped FOXP dimer (cyan and orange) bound to two separated DNA sites (top view).

(B) A side view of the model. The leucine zipper (LZ) preceding the forkhead domain, which may facilitate the formation of the domain-swapped dimer, is also shown as a cylinder. In this view, the backbone of the two strands of DNA are closer on the top of the arch-shaped dimer, where the protein has a relatively positive surface potential (see Figure 4D).

(C) Proposed roles of the FOXP dimer in DNA looping (left) and interchromosomal interaction (right).

By contrast, the Ala539Pro mutant of FOXP2 exists exclusively as a monomer in solution (Figure 4C). The mechanism by which this single amino acid change prevents swapping in classical FOX proteins appears to arise from proline's extraordinary propensity to disrupt α helices (Pace and Scholtz, 1998). Fortuitously, we have found that the optimal molar ratio of protein to DNA is 3:1 for crystallizing the FOXP2/DNA complex, allowing us to observe both the monomer- and domain-swapped dimer in the crystal.

The N termini of the FOXP2 forkhead domain in the swapped dimer are close to each other (Figure 3B). Interestingly, FOXP proteins contain a highly conserved zinc finger/leucine zipper motif about 50 residues N-terminal to the forkhead domain. This motif has been shown to mediate dimerization of FOXP proteins (Li et al., 2004; Wang et al., 2003). In the full-length protein, this zinc finger/leucine zipper may cooperate with the forkhead domain to facilitate the formation of domain-swapped dimers by FOXP proteins under physiological concentrations. However, we cannot rule out the possibility that the FOXP2 forkhead domain may also act as a monomer to bind DNA in vivo. The thermodynamics and functional implication of this monomer/dimer equilibrium by the FOXP2 forkhead domain remain to be investigated. The two H3 helices in the swapped FOXP2 dimer are separated sufficiently to allow both copies of FOXP2 to bind DNA simultaneously. However, because the two DNA binding surfaces are connected by a rigid protein domain characterized by extensive aromatic interactions (see above), the DNA binding sites of the domain-swapped FOXP2 dimer would need to be well-separated from each other or from separate DNA strands (Figure 5). Based on this structural feature, we propose that a unique function of the FOXP family of proteins is to loop DNA and/or mediate interchromosomal associations. Consistent with this proposed role, the convex surface is enriched in basic residues and has an overall positive electrostatic surface potential, which

may reduce phosphate backbone repulsion when two double-stranded DNAs are brought together (Figure 4D).

Disease-Related Mutations

FOXP2 and FOXP3 were discovered as the targets of genetic mutations in a human speech disorder (Lai et al., 2001) and the autoimmune disease IPEX (Bennett et al., 2001; Brunkow et al., 2001; Wildin et al., 2001), respectively. Although the potential effects of these mutations have been previously analyzed by homology modeling (Banerjee-Basu and Baxevanis, 2004), the accuracy of these analyses was limited by the divergence of the FOXP forkhead domain from the rest of the FOX family and the subtlety of many disease-linked mutations. The 1.9 Å resolution structures of FOXP2 and its complex with DNA described here allow us to examine these mutations in detail.

The mutation of Arg553 to histidine in human FOXP2 has been linked to a severe congenital speech disorder. Because of the relatively lower resolution of the previous FOXA3/DNA crystal structure, the precise role of this highly conserved arginine residue (Arg168 in FOXA3) in DNA recognition by the FOX family of transcription factors has not been clear (Clark et al., 1993). The present study shows that Arg553 is a major component of the FOXP2/DNA binding surface (Figure 2A). The guanidinium group of Arg553 forms a water-mediated hydrogen bond with Thy11′ while its long aliphatic side chain makes extensive van der Waals contacts to Thy11′, Ade12′, and Ade13′. Thus, the molecular mechanism of the congenital speech disorder is that the Arg553His mutation likely disrupts the organization of the protein/DNA interface in afflicted individuals.

Because of the high sequence identity between FOXP2 and FOXP3 (Figure 1C), the structure of FOXP2 can also be used to analyze the mutations in FOXP3 linked to human IPEX syndrome. These mutations include Ile363Val (Ile530 in FOXP2), Phe371Cys and Phe371Leu (Phe538 in FOXP2), Ala384Thr (Ala551 in FOXP2), and Arg397Trp (Arg564 in FOXP2). Ile530 is located on helix H2 (Figure 6A). Its Cδ carbon makes van der Waals contact with Leu527, Leu556, and Trp573. Leu527 and Trp573 contact DNA backbone ribose moieties directly, while Leu556 holds helix H3 in position to recognize DNA. The Ile363Val mutation appears to cause disease by a mechanism wherein it alters DNA binding by influencing neighboring residues that contact DNA. Thus, a relatively conservative mutation like Ile363Val in FOXP3 can potentially alter its DNA binding and ultimately lead to autoimmune disease. It is remarkable that a single methyl group, when placed strategically in protein, can have such significant impact on biology. This point is also well-illustrated in the Ala384Thr mutation (Figure 6B). In FOXP2, the corresponding residue Ala551 on helix H3 is intimately packed against Tyr509, while a serine residue, as seen in many classical FOX proteins, can be accommodated in this position. Thus, the Ala384Thr mutation in FOXP3 appears to cause disease by a mechanism wherein an extra γ-methyl group is forced into a tightly packed protein/DNA interface, thereby disrupting DNA binding. Arg564, located on strand S2, binds DNA on the backbone and in the minor groove (Figure 2A). Mutation of this arginine to the bulky and neutral tryptophan residue in FOXP3 (Arg397Trp) will likely destabilize the protein/DNA interaction through several effects including steric repulsion and the loss of the positive charge.

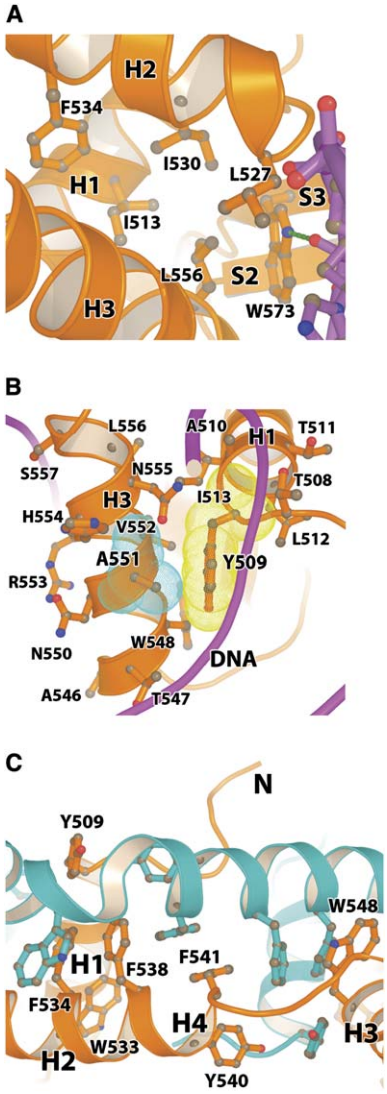

Figure 6. Disease Mutations

(A) IPEX mutation Ile363Val in FOXP3. The corresponding residue in FOXP2, Ile530, forms a cascade of van der Waals interactions with Leu527, Leu556, and Trp573 that contribute directly or indirectly to DNA binding.

(B) IPEX mutation Ala384Thr in FOXP3. The corresponding residue Ala551 in FOXP2 on helix H3 packs intimately against Tyr509. The Ala384Thr mutation would bring an extra γ-methyl group into a tightly packed protein/DNA interface and disrupt DNA binding.

(C) IPEX mutations Phe371Cys and Phe371Leu in FOXP3. The corresponding residue Phe538 in FOXP2 is critically located at the center of the domain-swapped dimer interface. Mutations of this phenylalanine residue to Cys or even a hydrophobic residue Leu may disrupt dimerization.

In mutations Phe371Cys and Phe371Leu, the corresponding residue Phe538 is exposed in the FOXP2 monomer structure and plays no apparent role in DNA binding and protein folding (Figure 3D). It is notable that FOXO4 has a valine at this position (Figure 1C), so a phenylalanine is not required for function here across the entire FOX family, arguing that this residue may have

a specialized function in the FOXP subfamily. Indeed, Phe538 is critically located at the center of the domain-swapped dimer interface (Figure 6C). Thus, mutations Phe371Cys and Phe371Leu in FOXP3 may exert their disease-causing effects by a mechanism wherein the core of the domain-swapped dimer is disrupted. Consistent with this observation, mutations disrupting the function of the leucine zipper motif in FOXP3 have also been linked to IPEX (Chatila et al., 2000). Overall, the high-resolution crystal structure of the FOXP2/DNA complex provides a structural basis for understanding disease-causing mutations in FOXP2 and FOXP3, which in turn support the physiological relevance of our structural and biochemical observations.

In summary, high-resolution structural studies of the FOXP2/DNA complex reported here have revealed the general mechanism of DNA recognition by the FOX family of transcription factors and unusual biochemical properties of the FOXP2 forkhead domain. The most surprising finding from these studies is that the FOXP2 forkhead domain can form a domain-swapped dimer. Disease-related mutations, sequence comparison, and biochemical analyses argue strongly that this domain swapping is a physiologically relevant function evolved in the P branch of FOX proteins. Though the biological significance of swapping in FOXP2 has yet to be determined, swapping is thought to have functional roles in several other proteins. The most notable examples are found in the cadherins, which utilize an exchange of N-terminal β strands to mediate dimerization in cell-cell adhesion (Chen et al., 2005; Shapiro et al., 1995). Unlike most dimeric transcription factors that bind adjacent DNA sites, modeling predicts that domain-swapped FOXP dimers can only bind cognate DNA sites separated far from each other or located on different DNA strands, suggesting that a specialized function of FOXP proteins in transcriptional regulation is to promote the assembly of higher order protein/DNA complexes. These insights will facilitate further studies of the role of FOXP2 in human language development and FOXP3 in immunological tolerance.

Experimental Procedures

Sample Preparation and Crystallization

The forkhead domain of human FOXP2 (amino acids 503–584) was cloned into the pET-30 LIC vector as a 6× histidine-tagged fusion protein and was expressed in Rosetta pLysS cells (Novagen, San Diego, CA). The protein was first purified by Ni-NTA beads (Qiagen, Valencia, CA) and then digested by enterokinase to remove the histidine tag. The digested forkhead protein was further purified by Mono S cation exchange and Superdex 75 size exclusion column (Amersham Biosciences, Piscataway, NJ). Although the protein could be purified as separated dimer and monomer by size exclusion column, it reequilibrates as a monomer/dimer mixture after 24 hr at room temperature or 10 hr at 37ºC. We therefore did not attempt to separate the two species for crystallization. The protein (both monomer and dimer) was then concentrated to approximately 8 mg/ml in 5 mM HEPES (pH 7.63), 2 mM β-mercaptoethanol, 0.5 mM EDTA, and 150 mM NaCl (S75 buffer) and stored at 4ºC. DNA was synthesized by Integrated DNA Technologies (Coralville, IA) and purified as described previously (Chen et al., 1998). The two strands of DNA are 5′-AACTATGAAACAAATTTTCCT-3′ and 5′-TTAGGAAAA TTTGTTTCATAG-3′. The protein/DNA complex was prepared by mixing protein and DNA at 3:1 molar ratio in S75 buffer at a final concentration of 10 mg/ml. Crystals were grown by the hanging drop method at 18ºC using a reservoir buffer of 50 mM bis-tris-propane (BTP) (pH 6.68), 100 mM NaCl, 6% (w/v) PEG 3000, 10 mM $MgCl_2$, and 0.01% (w/v) sodium azide. Typically, crystals grew to approximately 100 × 100 × 50 μm in 1–2 weeks. Crystals belong to the space group $P2_1$ with cell dimensions a = 67.54 Å, b = 124.21 Å, c = 67.67 Å, and β = 110.81º.

Data Collection, Structure Determination, and Analysis

Crystals were stabilized in the harvest/cryoprotectant buffer: 50 mM BTP (pH 6.68), 100 mM NaCl, 21% (w/v) PEG 3000, and 25% (w/v) glycerol and flash frozen with liquid nitrogen for cryocrystallography. Data were collected at the ALS BL8.2.1 beamline at the Lawrence Berkeley National Laboratory and BioCARS sector 14-BM-C of Advanced Photon Source, Argonne National Laboratory. Data were reduced using DENZO and SCALEPACK (Otwinowski and Minor, 1997). Initial phases were determined by molecular replacement using the coordinates of FOXA3 (Clark et al., 1993). Molecular replacement, refinement, and final analysis were done with CNS (Brunger et al., 1998). Final models have very good geometry. All residues are in the allowed regions of the Ramachandran plot, with 92% in the most favored regions. The statistics of crystallographic analysis are presented in Table 1. Figures of structure illustration were prepared using MOLSCRIPT (Kraulis, 1991) and Pymol (DeLano Scientific, San Francisco, CA). Model building and structural comparisons were carried out in O (Jones et al., 1991).

Mutagenesis and Multiangle Light Scattering Analysis

The Ala539Pro mutation in human FOXP2 (amino acids 503–584) was made by site-specific mutagenesis (QuikChange; Stratagene, La Jolla, CA). The mutant was expressed and purified similarly to the wild-type protein (see above). Both the mutant and the wild-type protein were analyzed by a multiangle light scattering detector (DAWN-EOS; Wyatt Technologies, Santa Barbara, CA) after separation by a size exclusion column (KW-803; Shodex, Aston, PA) on HPLC.

Acknowledgments

The authors thank Liang Guo, Luke G. Kroiss, Andrew Bonham, Greg Smith, and Keith Brister for help in data collection and discussions; Youngchang Kim, James Whisstock, Ning Lei, Pearl Quartey, and Frank Collart for assistance in initial protein expression, purification, and bioinformatics analysis; and Jim Goodrich, Marcelo Sousa, Debbie Wuttke, and Mitchell Guss for critical reading of the manuscript. H.T. was supported by the Australian Synchrotron Research Program (ASRP). The use of the BioCARS sector at the Advanced Photon Source was supported by ASRP, NIH/National Center for Research Resources, and the U.S. Department of Energy. This research is supported by grants from the W.M. Keck Foundation (L.C.) and NIH (L.C.).

Received: August 3, 2005
Revised: September 27, 2005
Accepted: October 4, 2005
Published: January 10, 2006

References

Banerjee-Basu, S., and Baxevanis, A.D. (2004). Structural analysis of disease-causing mutations in the P-subfamily of forkhead transcription factors. Proteins *54*, 639–647.

Bennett, C.L., Christie, J., Ramsdell, F., Brunkow, M.E., Ferguson, P.J., Whitesell, L., Kelly, T.E., Saulsbury, F.T., Chance, P.F., and Ochs, H.D. (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat. Genet. *27*, 20–21.

Bettelli, E., Dastrange, M., and Oukka, M. (2005). Foxp3 interacts with nuclear factor of activated T cells and NF-κ B to repress cytokine gene expression and effector functions of T helper cells. Proc. Natl. Acad. Sci. USA *102*, 5138–5143.

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., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Brunkow, M.E., Jeffery, E.W., Hjerrild, K.A., Paeper, B., Clark, L.B., Yasayko, S.A., Wilkinson, J.E., Galas, D., Ziegler, S.F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat. Genet. *27*, 68–73.

Carlsson, P., and Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. Dev. Biol. *250*, 1–23.

Chatila, T.A., Blaeser, F., Ho, N., Lederman, H.M., Voulgaropoulos, C., Helms, C., and Bowcock, A.M. (2000). JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. J. Clin. Invest. *106*, R75–R81.

Chen, L., Glover, J.N., Hogan, P.G., Rao, A., and Harrison, S.C. (1998). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. Nature *392*, 42–48.

Chen, C.P., Posy, S., Ben-Shaul, A., Shapiro, L., and Honig, B.H. (2005). Specificity of cell-cell adhesion by classical cadherins: critical role for low-affinity dimerization through β-strand swapping. Proc. Natl. Acad. Sci. USA *102*, 8531–8536.

Clark, K.L., Halay, E.D., Lai, E., and Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature *364*, 412–420.

Clifton-Bligh, R.J., Wentworth, J.M., Heinz, P., Crisp, M.S., John, R., Lazarus, J.H., Ludgate, M., and Chatterjee, V.K. (1998). Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nat. Genet. *19*, 399–401.

Costa, R.H., Grayson, D.R., and Darnell, J.E., Jr. (1989). Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and α 1-antitrypsin genes. Mol. Cell. Biol. *9*, 1415–1425.

Fang, J., Dagenais, S.L., Erickson, R.P., Arlt, M.F., Glynn, M.W., Gorski, J.L., Seaver, L.H., and Glover, T.W. (2000). Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome. Am. J. Hum. Genet. *67*, 1382–1388.

Finegold, D.N., Kimak, M.A., Lawrence, E.C., Levinson, K.L., Cherniske, E.M., Pober, B.R., Dunlap, J.W., and Ferrell, R.E. (2001). Truncating mutations in FOXC2 cause multiple lymphedema syndromes. Hum. Mol. Genet. *10*, 1185–1189.

Frank, J., Pignata, C., Panteleyev, A.A., Prowse, D.M., Baden, H., Weiner, L., Gaetaniello, L., Ahmad, W., Pozzi, N., Cserhalmi-Friedman, P.B., et al. (1999). Exposing the human nude phenotype. Nature *398*, 473–474.

Jin, C., Marsden, I., Chen, X., and Liao, X. (1999). Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. J. Mol. Biol. *289*, 683–690.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519–523.

Lehmann, O.J., Ebenezer, N.D., Jordan, T., Fox, M., Ocaka, L., Payne, A., Leroy, B.P., Clark, B.J., Hitchings, R.A., Povey, S., et al. (2000). Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma. Am. J. Hum. Genet. *67*, 1129–1135.

Lehmann, O.J., Sowden, J.C., Carlsson, P., Jordan, T., and Bhattacharya, S.S. (2003). Fox's in development and disease. Trends Genet. *19*, 339–344.

Levy-Lahad, E., and Wildin, R.S. (2001). Neonatal diabetes mellitus, enteropathy, thrombocytopenia, and endocrinopathy: further evidence for an X-linked lethal syndrome. J. Pediatr. *138*, 577–580.

Li, C., and Tucker, P.W. (1993). DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/fork head domain. Proc. Natl. Acad. Sci. USA *90*, 11583–11587.

Li, S., Weidenfeld, J., and Morrisey, E.E. (2004). Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol. Cell. Biol. *24*, 809–822.

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). Solution structure of the DNA-binding domain of interleukin enhancer binding factor 1 (FOXK1a). Proteins *49*, 543–553.

Mazet, F., Yu, J.K., Liberles, D.A., Holland, L.Z., and Shimeld, S.M. (2003). Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. Gene *316*, 79–89.

Mears, A.J., Jordan, T., Mirzayans, F., Dubois, S., Kume, T., Parlee, M., Ritch, R., Koop, B., Kuo, W.L., Collins, C., et al. (1998). Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly. Am. J. Hum. Genet. *63*, 1316–1328.

Mirzayans, F., Gould, D.B., Heon, E., Billingsley, G.D., Cheung, J.C., Mears, A.J., and Walter, M.A. (2000). Axenfeld-Rieger syndrome resulting from mutation of the FKHL7 gene on chromosome 6p25. Eur. J. Hum. Genet. *8*, 71–74.

Nirula, A., Moore, D.J., and Gaynor, R.B. (1997). Constitutive binding of the transcription factor interleukin-2 (IL-2) enhancer binding factor to the IL-2 promoter. J. Biol. Chem. *272*, 7736–7745.

Nishimura, D.Y., Swiderski, R.E., Alward, W.L., Searby, C.C., Patil, S.R., Bennet, S.R., Kanis, A.B., Gastier, J.M., Stone, E.M., and Sheffield, V.C. (1998). The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. Nat. Genet. *19*, 140–147.

Nishimura, D.Y., Searby, C.C., Alward, W.L., Walton, D., Craig, J.E., Mackey, D.A., Kawase, K., Kanis, A.B., Patil, S.R., Stone, E.M., and Sheffield, V.C. (2001). A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. Am. J. Hum. Genet. *68*, 364–372.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, C.W.J. Carter and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.

Pace, C.N., and Scholtz, J.M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. Biophys. J. *75*, 422–427.

Schubert, L.A., Jeffery, E., Zhang, Y., Ramsdell, F., and Ziegler, S.F. (2001). Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. J. Biol. Chem. *276*, 37672–37679.

Shapiro, L., Fannon, A.M., Kwong, P.D., Thompson, A., Lehmann, M.S., Grubel, G., Legrand, J.F., Als-Nielsen, J., Colman, D.R., and Hendrickson, W.A. (1995). Structural basis of cell-cell adhesion by cadherins. Nature *374*, 327–337.

Shu, W., Yang, H., Zhang, L., Lu, M.M., and Morrisey, E.E. (2001). Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J. Biol. Chem. *276*, 27488–27497.

van Dongen, M.J., Cederberg, A., Carlsson, P., Enerback, S., and Wikstrom, M. (2000). Solution structure and dynamics of the DNA-binding domain of the adipocyte-transcription factor FREAC-11. J. Mol. Biol. *296*, 351–359.

Wang, B., Lin, D., Li, C., and Tucker, P. (2003). Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J. Biol. Chem. *278*, 24259–24268.

Weigelt, J., Climent, I., Dahlman-Wright, K., and Wikstrom, M. (2001). Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). Biochemistry *40*, 5861–5869.

Wildin, R.S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J.L., Buist, N., Levy-Lahad, E., Mazzella, M., Goulet, O., Perroni, L., et al. (2001). X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat. Genet. *27*, 18–20.

Accession Numbers

Coordinates have been deposited in the RCSB Protein Data Bank under accession code 2A07.