University of Alberta

Functional analysis of the wing 2 sub-domain and sequence recognition properties of the FOXC1 forkhead domain

By

Tara Camille Murphy

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

In

Medical Sciences – Ophthalmology

Edmonton, Alberta
Fall 2004

Your file Votre référence
ISBN: 0-612-95824-8
Our file Notre référence
ISBN: 0-612-95824-8

Acknowledgements

First and foremost I would like to thank Dr. Michael Walter for giving me the opportunity to discover science and providing a stellar lab to work in. I would like to say a huge thank-you to Dr. Fred Berry for allowing me one question per day and always providing me with an answer or helpful discussion. Thanks to Matthew and Sharp for being the big brothers I never had. The warmest thanks to Farideh Mirzayans who, without, we would all be lost. The rest of the Walter lab, I thank you for all your help! Thanks to Kelly, for sneaking me sundaes and being an honorary member of our lab. I would like to thank the members of my committee, Dr. Alan Underhill, Dr. Moira Glerum, Dr. Roseline Godbout, and Dr. Susan Andrew. To my mom and dad, thank-you for your unwavering support and interest, I love you both. Thanks to everyone for their love and support!

Table of Contents

List Of Tables

CHAPTER ONE

CHAPTER TWO

List of Figures

INTRODUCTION

CHAPTER ONE

indicated

CHAPTER TWO

List of Abbreviations

AR – Axenfeld-Rieger malformations

bp – base pairs

ch – congenital hydrocephalus

DAPI – 4,6-diamidino-2-phenly-indole

dpc – days post coitum

EMSA – electrophoretic mobility shift assay

FHD- forkhead domain

FOX- Forkhead box

HTH- helix turn helix

IH – iris hypoplasia

IGD - iridogoniodysgenesis

kDa – kilodaltons

kb - kilobase

L-dopa – L-dihydroxyphenylalanine

N - nuclear

N + C – nuclear and cytoplasmic

NLS – nuclear localization signal

NMR- nuclear magnetic resonance

PSM – presomitic mesoderm

SC- Schlemm's canal

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

TD – transactivation domains

TK- thymidine kinase

TM – trabecular meshwork

Tyr- Tyrosinase

wHTH- winged helix turn helix

Introduction

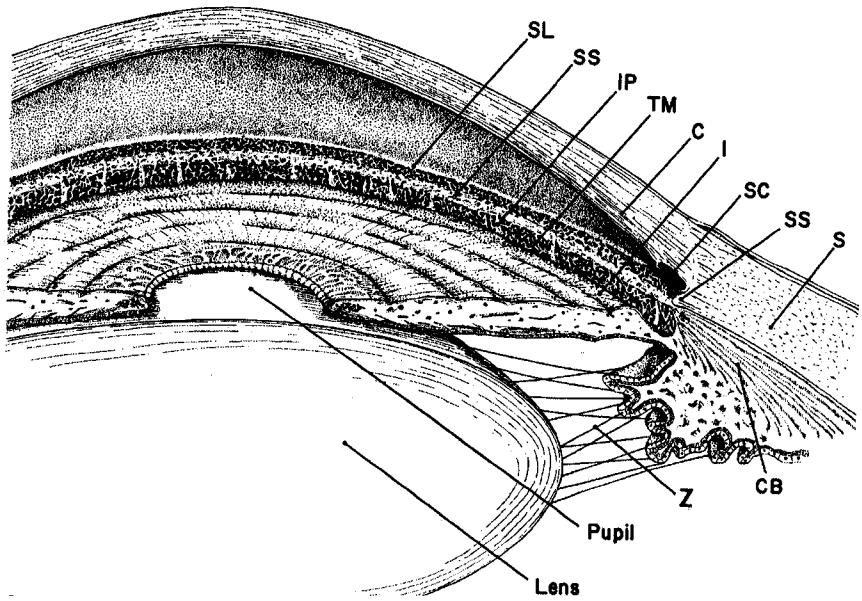
Incidence and clinical description of glaucoma

Glaucoma is the second leading cause of blindness in the world, after cataract, affecting 67 million people worldwide (1). Glaucoma is characterized by progressive damage to the optic nerve and usually begins with loss of peripheral vision, with loss of central vision occurring in the later stages (2). Because the beginning of vision loss is subtle, many cases of glaucoma go undetected until optic nerve damage has advanced. Early detection can prevent nerve damage and early stages of glaucoma can be treated with surgery and drugs (2). Glaucoma results from both hereditary and non-hereditary causes, with the former being divided into primary, secondary and congenital hereditary glaucoma. These divisions in turn, can be further sub-divided into open or closed angle glaucoma, which is determined by clinical examination of the irido-corneal angle.

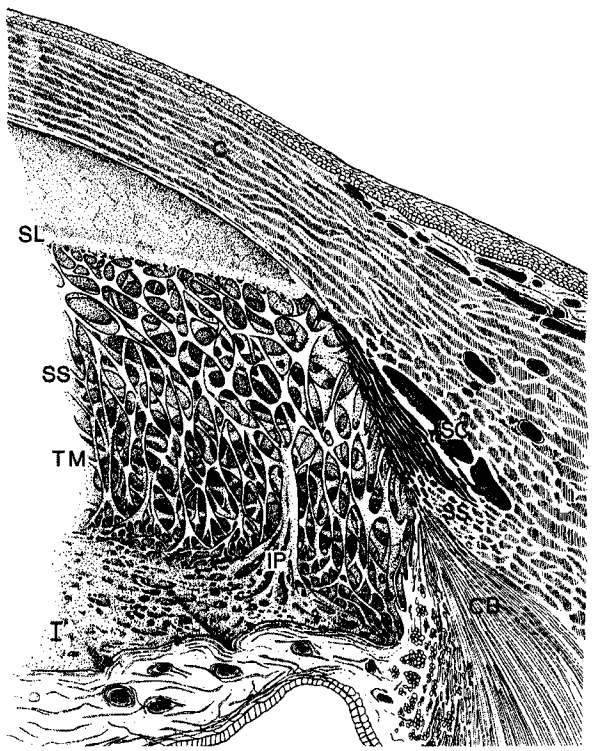
Structure and function of the anterior chamber/angle of the eye

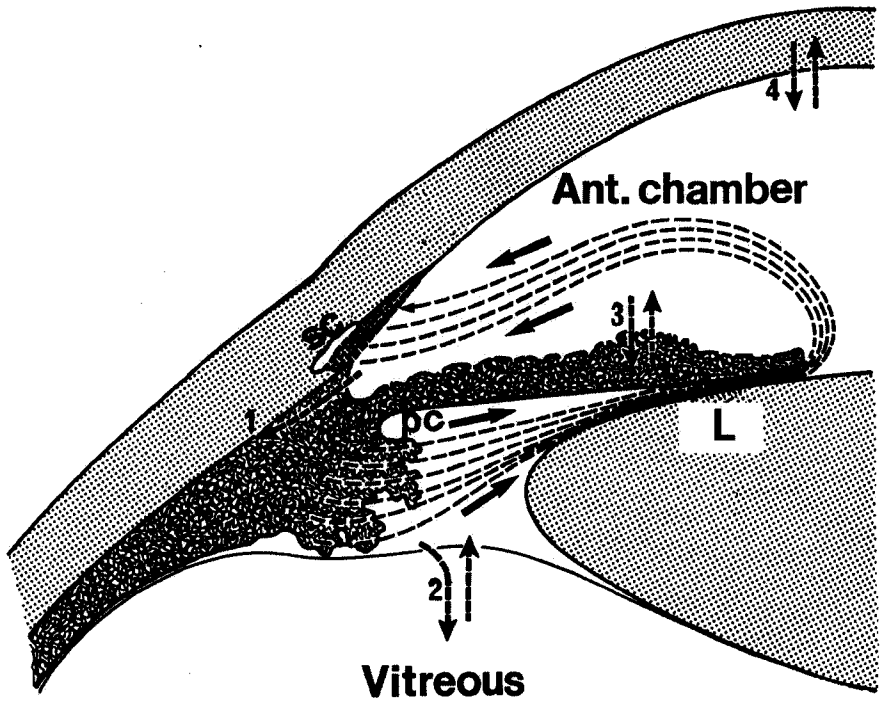
The chamber angle is formed by the juxtaposition of the root of the iris and the peripheral cornea, hence the term *irido-corneal angle* (2) (figure i-1). The significance of the chamber angle lies in the fact that the main drainage pathway of the aqueous humor is located in this region. Aqueous humor is formed by the ciliary process and flows into the anterior chamber through the pupil. The main drainage route for the bulk outflow is constituted by the trabecular meshwork (TM) and Schlemm's canal (SC) (3). The aqueous humor flows through the TM into SC, the collector channels of SC, and then into the intrascleral and episcleral venus plexuses (4) (figure i-2). The TM is composed of non-filtering and filtering regions and one of its crucial functions is to 'clean' the aqueous humor, preventing obstruction (4). In a normal eye, the angle offers a steady-state resistance to fluid outflow in response to inflow of aqueous humor. The pressure of inflow rises to a level sufficient to drive fluid across that resistance at the same rate at which the ciliary body produces aqueous humor (4).

A



B





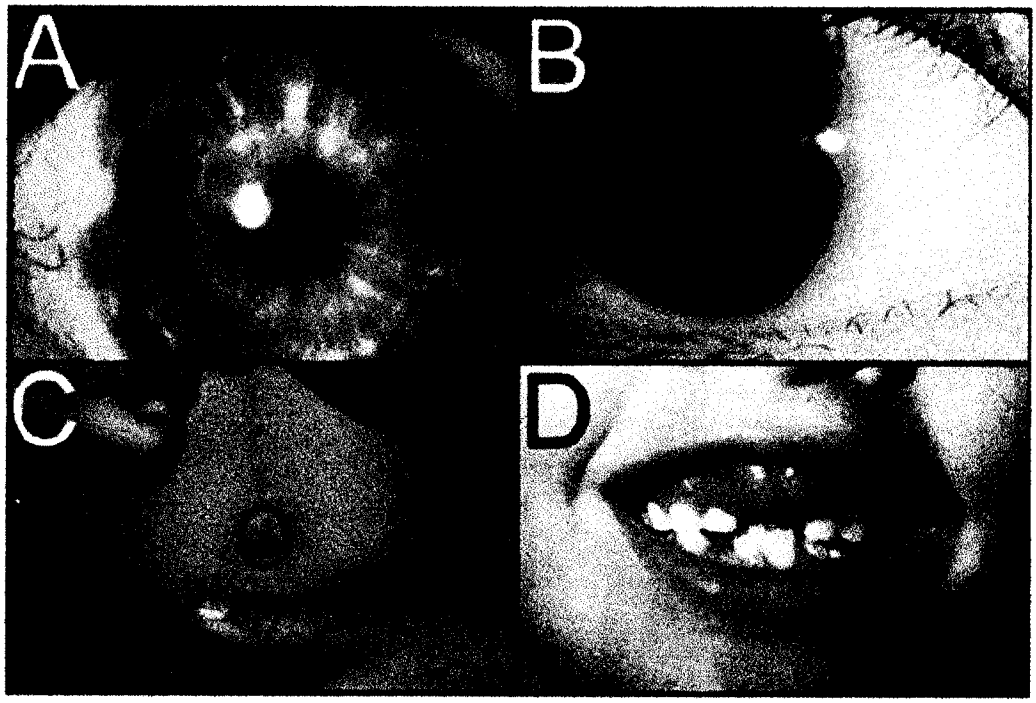
With many congenital glaucomas, malformed drainage structures or increase in resistance of outflow are thought to increase intraocular pressure and exert stress on the optic nerve head, eventually causing blindness.

Clinical description of Axenfeld-Rieger malformations

Axenfeld-Rieger syndrome is a clinically heterogeneous condition with variable expressivity. In 1920, Axenfeld diagnosed a condition which he called Axenfeld *anomaly* (5), where the iris strands bridge the iridocorneal angle to the trabecular meshwork and posterior embryotoxon is present (prominent and anteriorly displaced Schwalbe line) (figure i-3 a and b)(6). Later, in 1935 Rieger (7) diagnosed ocular malformations, which he recognized as inherited, as Rieger syndrome and Rieger anomaly. Initially, Rieger *anomaly* was diagnosed if a patient presented with stromal hypoplasia of the iris, distorted or displaced pupils (corectopia), or extra holes in the iris (polycoria) (6). Rieger *syndrome* was diagnosed if the patient presented with Rieger anomaly in addition to systemic phenotypes such as midface abnormalities (maxillary hypoplasia, telecanthus, hypertelorism, and broad, flat, nasal bridge), dental anomalies (hypodontia and microdontia), and/or redundant umbilical skin (6, 8) (figure i-3 c and d). Ocular and systemic defects overlap between Axenfeld anomaly, Rieger anomaly and Rieger syndrome, making the separate classifications redundant and somewhat arbitrary. As molecular data begins to shape our understanding of the pathogenesis of Axenfeld anomaly and Rieger anomaly and syndrome, the term Axenfeld-Rieger malformations is more apt as an encompassing term to describe the overlapping defects.

Tissues affected in AR malformations share neural crest origin

In studying any disease that develops congenitally or presents with early onset, it is crucial to understand the embryonic origins of the tissues affected. With AR malformations, it is thought that both the ocular and systemic findings result from abnormal migration or defective terminal induction of neural crest cells (9) (10). Neural crest cells are those neuroectodermal cells that proliferate from the crest of the neural folds when the folds fuse to form the neural tube in the embryo (9).



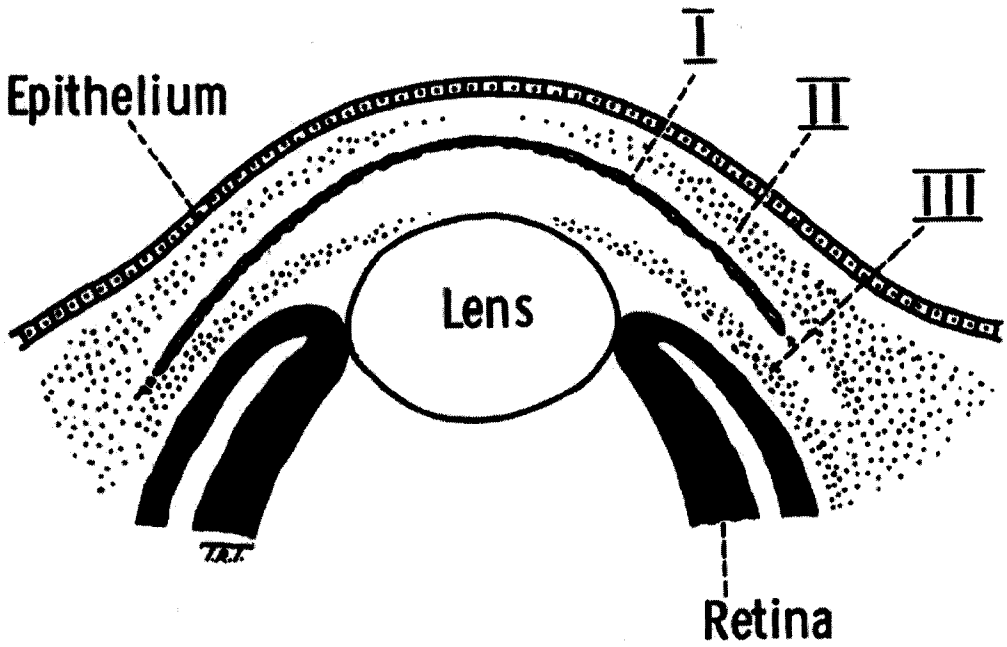
Neuronal-specific enolase is the gamma-gamma form of the glycolytic enzyme enolase and its expression in normal cells is believed to indicate cell differentiation from neuroectoderm (9). Positive reaction to enolase antibodies has been used as a method to determine the embryonic origin of structures affected in ocular diseases. Cells from the neural crest migrate into the developing ocular areas in three successive waves (3). The first wave forms the corneal endothelium when neural crest cells invade the presumptive eye and form an endothelial layer on the anterior chamber. The second wave forms the corneal stroma, uveal meshwork and pupillary membrane. This is followed by a final wave of neural crest cells, which form the iris stroma (3) (figure i-4). The development and differentiation of the corneal endothelium occurs in conjunction with the development of the structures of the eye involved in aqueous outflow, namely the trabecular meshwork and Schlemm's canal. A predominant theory reasons that blockage or functional abnormalities that lead to increased resistance to outflow can lead to an increase in intraocular pressure which can result in optic nerve damage, eventually leading to glaucoma (12).

Genetic loci involved in AR malformations

PITX2 (pituitary homeobox transcription factor 2)

PITX2 (RIEG1) was positionally cloned by Semina *et al. (13)* and is located on chromosome 4 at locus q25. PITX2 is a member of the paired-bicoid family of homeodomain transcription factors. Numerous pathologic mutations have been found in PITX2 that produce a spectrum of clinical phenotypes which include AR malformations, IGD, and IH (13) (14) (15) (16) (17). Characterization of patient mutations suggest that PITX2 proteins that retain partial function result in milder anterior segment dysgenesis, implying a correlation between genotype and phenotype (18).

Pitx2 mouse models have been invaluable in gaining insight into the pathogenesis of human *PITX2* mutations. Murine *Pitx2* shares a similar expression pattern to human *PITX2*, with expression in the periocular mesenchyme, umbilicus, heart, gut, dental epithelia and limb bud (19) but is undetected in the optic stalk and eminence (20). Homozygosity for null or hypomorphic alleles cause death by day 10



of gestation in part due to incomplete neural tube closure, lung isomerism and failed heart septation (21). Through targeted gene disruption, a series of *Pitx2* hypomorphic and null alleles were created which resulted in a plethora of murine phenotypes that mimic human AR malformations, providing a useful animal model (21).

FOXC1 (forkhead box transcription factor C1)

FOXC1 (formerly FREAC3 and FKHL7) is a member of the FOX (Forkhead Box) family of transcription factors and mutations in *FOXC1* underlie Axenfeld-Rieger malformations (22). *FOXC1* contains a conserved DNA-binding domain known as the forkhead domain (FHD). This 110 amino acid, monomeric binding domain was first identified with the discovery of the *Drosophilia melanogaster* protein *Fork head* and the rat Hepatocyte Nuclear Factor 3 (HNF3) proteins (23, 24). FOX family members have proven to be important in a multitude of developmental processes such as embryogenesis, cell cycle regulation, cell signaling cascades, and tissue-specific cell differentiation (25).

Located at p25 on human chromosome 6, *FOXC1* is a single exon gene that produces a protein of 553 amino acids, from an open reading frame of 1659 bp. Northern blot analysis shows that a 4.5 kb FOXC1 mRNA has variable expression over a range of adult and fetal tissues with highest expression in adult heart, pancreas, peripheral blood leukocytes, kidney and fetal kidney (26). An alternative 4.0 kb transcript is also present in fetal kidney suggesting use of an alternative promoter or polyadenylation site in this tissue (26).

Foxc1 (formerly *Mf1 and Fkh1)* is the murine homolog of human *FOXC1*. *Foxc1* mRNA is expressed in all adult organs except in liver, with highest expression found in heart, kidney, adrenals and brain (27). The human and mouse sequences show 97% identity overall with 100% identity in the forkhead domain itself.

Cardiac expression of FOXC1

In the murine embryo, *Foxc1* expression is seen in the developing heart as early as 7.5 days post coitum (dpc) (27) and seems to correlate with the formation of valves and atrial septae of the murine heart. $Foxc1^{LacZ}$ expression is detected

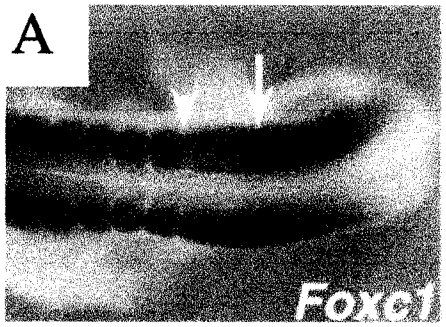
throughout the pharyngeal arch system and weakly in the endothelium of the heart at 9.5 dpc and by 10.5 dpc expression is seen in the mesenchyme and endothelium of the aortic arches (28). By 13.5 dpc *Foxc1* is expressed in the ventral part of the septum primum, which forms the atrial septum. Expression is also seen in the venous valve, tricuspid valve, and the mitral valve (29), where expression is maintained in these regions until 15 dpc (28).

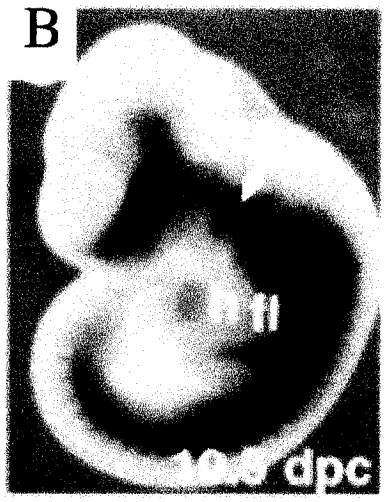
Urogenital expression of FOXC1

At 8.5 dpc, *Foxc1* is present in the presumptive intermediate mesoderm as well as the presomitic mesoderm and somites (30). Expression continues into 9.5 dpc when the nephrogenic cord develops and the Wolffian (nephric) duct begins to elongate caudally along the embryo in the intermediate mesenchyme (30). At 9.5 dpc *Foxc1* expression is intense in the mesonephric mesenchyme alongside the Wolffian duct followed by later expression in the mesonephric tubules (31). When formation of the metanephric kidney begins at 10.5 dpc, *Foxc1* expression is seen in the metanephric mesenchyme but not in the epithelium of the Wolffian duct or branching ureter (30). By 12.5 dpc transcript expression is still detected in the condensing mesenchyme of the kidney but at reduced levels.

Foxc1 expression in the somites and presomitic mesoderm

Foxc1 expression is first seen at 7.5-9.5 dpc and is restricted to the entire non-notochord mesoderm, with highest expression levels in the paraxial regions (32, 33). Expression is seen in two continuous wide bands running parallel to the neural plate/tube and primitive streak and extend from the presomitic mesoderm to the head mesoderm (33) (figure i-5). As somitogenesis proceeds, it becomes evident that *Foxc1* is required for skeletal development, where it is expressed by 11.5-12.5 dpc in the condensing mesenchyme of the vertebrae and forelimbs (31, 34).





Ocular expression of FOXC1

Foxc1 expression is seen in developing ocular tissues beginning around 11.5 dpc. (36) At this time, *foxc1* is expressed in the presumptive corneal mesenchyme cells between the corneal epithelium and the lens (31). As well, expression is seen in mesenchyme cells in the optic cup between the lens and retina; however, *foxc1* expression is not seen in the lens. An extensive area of periocular mesenchyme surrounding the eye, including the prospective trabecular meshwork and sclera, as well as the future conjunctival epithelium express *Foxc1* (36). After 16.5 dpc, when the eyelids are closed, *Foxc1* expression is restricted to the prospective trabecular meshwork cells, the sclera and the conjunctival epithelium (36).

Other Loci: Leucine zipper transcription factor MAF

Linkage analysis and chromosomal rearrangements suggest there are additional AR loci at 13q14 and 16q24 that remain to be cloned (37). The locus 13q14 is linked to AR malformations in conjunction with sensorineural hearing loss (38) and may confer a clinically distinct disorder. Candidate gene searches suggested that *FOXC2* might be a logical candidate based on its proximal location on 16q, its significant homology to *FOXC1*, and the phenotype of the double heterozygote *FOXC1* +/- *FOXC2* +/- mouse (refer to page 20). However, mutations in *FOXC2* associate with lymphedema-distichiasis rather than resulting in AR malformations (37, 39). Recently however, detailed ocular phenotyping of patients from nine different *FOXC2* pedigrees revealed that mutations in *FOXC2* can result in anterior segment anomalies and corneal thickening, suggesting that the role *FOXC2* plays in AR malformations is not fully defined (40). Ocular abnormalities (cataract, anterior segment dysgenesis and microphthalmia) were found to co-segregate with a translocation t(5;16)(p15.5;q23.2) (41). Cloning of the breakpoint lead to the discovery of the basic leucine zipper transcription factor MAF. MAF is expressed in the vertebrate lens and is involved in lens development and expression of lens crystallins. Screening of families has shown a link between mutations in MAF and defects in lens development and associated anterior segment abnormalities (41). As well, many anterior segment disorders are thought to arise from abnormal migration

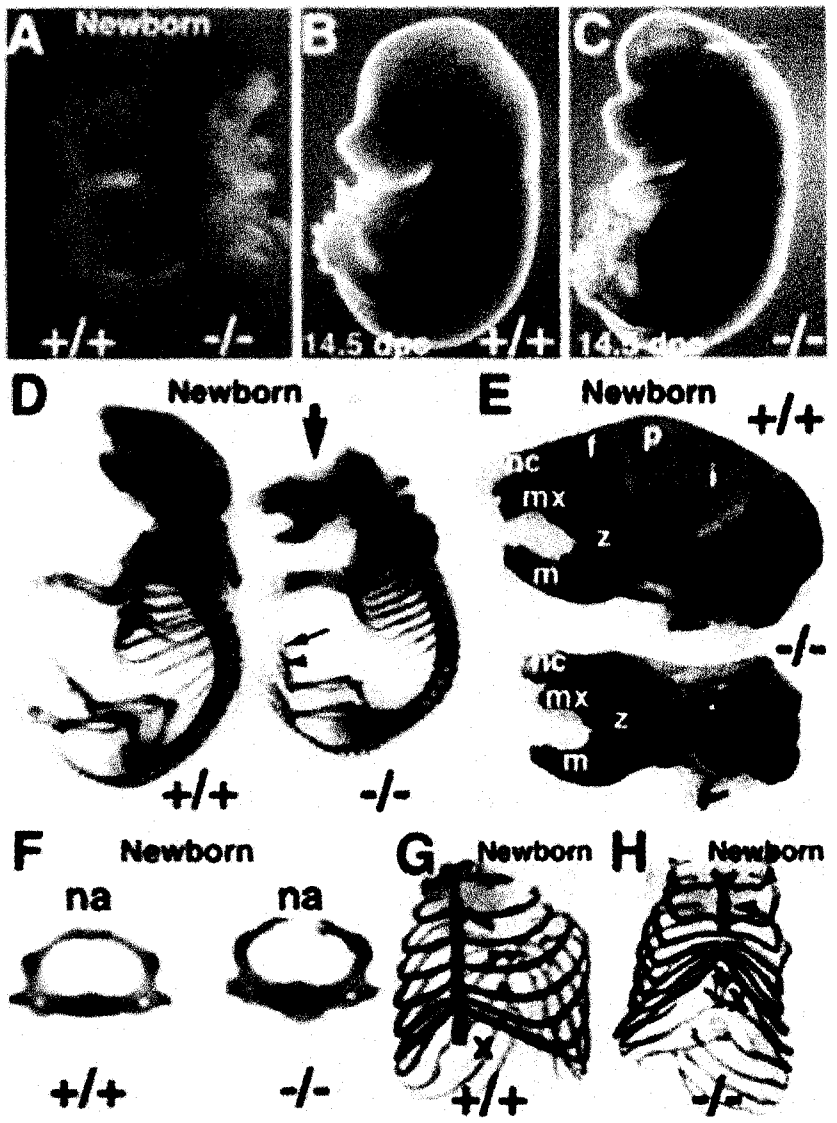
of neural crest cells (see page 6,9). The lens and anterior segment are linked in development through a necessity of separation of these components for anterior chamber formation and a requirement for signaling from the lens for anterior segment organization (42). It is conceivable then, that while some anterior segment disorders result from abnormal neural crest migration, others arise due to abnormalities in lens development (41). Therefore, the location of MAF at 16q23.2, and its co-segregation with defects in anterior segment anomalies, make MAF a good candidate gene for causing AR malformations.

Foxc1 mouse models

Foxc1 mouse models have been used extensively to help delineate *Foxc1* expression and gain understanding into *foxc1*'s role in development. The spontaneous *congenital hydrocephalus (ch)* mouse mutation was first identified by Gruneberg in 1943. Over fifty years later, homozygosity for a nonsense mutation in *Foxc1* was found, which produced a truncated form of the protein (lacking the DNA-binding domain) that resulted in *ch (31)*. The *ch* phenotype is identical to the homozygous $Foxc1^{Lacz}$ null mutant phenotype *(31)*.

ch mice have a variety of developmental defects

Ch mice have a congenital lethal hydrocephalus in association with multiple developmental defects affecting the central nervous, urogenital and skeletal systems (figure i-6) (43). Before 10.5 dpc *Foxc1* -/- cannot be distinguished from their wild type littermates, but one day later there is a clear difference in brain size, which becomes more discernable by 13.5 dpc (31, 43). *Ch* mice develop a steeply bulging forehead, corresponding to the cerebral hemispheres, with grossly hemorrhagic cerebral spinal fluid in the massively dilated ventricular system (43), which manifests as a dark purplish-blue color. Most affected mice survive to term but shortly after they take their first breath, die of asphyxia (36, 43). All mutants exhibit craniofacial defects with the most striking being the absence of bony calvarium in the hydrocephalus due to absence of frontal, parietal and interparietal bones (36, 43) (figure i-6 d-e). The basioccipital, exoccipital and hyoid bones are smaller and



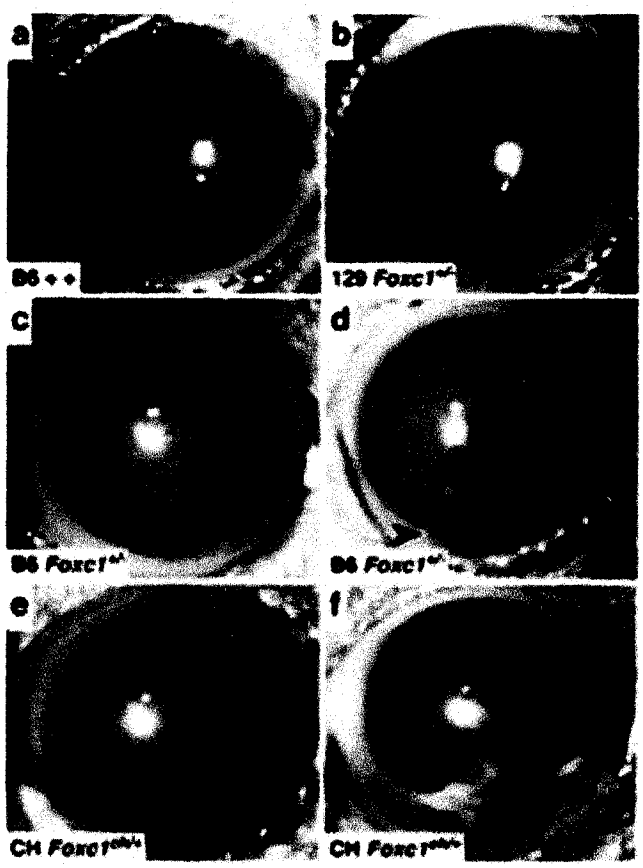
malformed while the squamosal and zygomatic bones are severely ossified (36, 43). The nasal septum is shorter and reduced giving the mutants a characteristic short snout appearance (36, 43). In addition, the rib cages of *ch* mice are extremely fragile due to lack of a sternum and poorly attached left and right costal cartilages (36, 43) (figure 6 f-h). Finally, in the appendicular skeleton the digits are thinner than normal with smaller ossification centers (31).

$Foxc1^{lacZ}$ and ch mice have ocular defects similar to Axenfeld-Rieger malformations

$Foxc1^{Lacz}$ or ch heterozygotes have multiple anterior segment abnormalities with variable expressivity that resemble those of human AR malformations (43, 20). However, the penetrance of Foxc1 defects is low in *$Foxc1^{Lacz}$* mice. The most common phenotype is (bilateral) corectopia and irregularly shaped pupils in addition to posterior embryotoxon and corneal opacity (43) (figure i-7). Corectopia progresses to iridocorneal adhesion and polycoria (multiple holes in the iris) becomes evident. In a wild type mouse, the anterior chamber forms as a result of separation between the cornea and the lens around 13.5-14.5 dpc (36). In the *ch* -/- mouse this separation does not take place and an anterior chamber does not develop. Abnormalities with the irido-corneal angle occur, such as smaller or absent Schlem's canal, large blood vessels and iris strands, and hypoplastic or absent trabecular meshwork (44). By 16.5 dpc the eyelids have closed in the wild type mouse, unlike the *ch* -/- mouse, where the eyelids remain open (44).

Foxc1 and Foxc2 operate in the same developmental systems

Foxc2 is highly related to Foxc1, sharing 97% identity (99% similarity) in their DNA-binding domains (30). *Foxc1* and *Foxc2* have overlapping expression in a number of developing systems, namely skeletal, urogenital, ocular and cardiac tissues (28, 30, 35). Both Foxc1 +/- and Foxc2 +/- single heterozygotes die pre- or perinatally from multiple cardiac and skeletal defects (28, 44). The spectrum of phenotypes in the urogenital, ocular and cardiac systems seen with Foxc1+/- and Foxc2 +/- single heterozygotes are recapitulated in the Foxc1+/- / Foxc2+/-



compound heterozygote, demonstrating that to some extent, both genes are operating in the same developmental systems. Compound homozygotes die earlier, with more severe cardiac and somitogenesis defects than single homozygotes. Compound homozygotes show complete absence of segmental paraxial mesoderm and abnormalities in the branchial arches and remodeling of the blood vessels (35). The phenotype of the homozygote null embryos and the nonallelic complementation of the two null mutations suggest that Foxc1 and Foxc2 play similar, dose-dependent roles in cardiovascular development and somitogenesis (28). *Foxc1* and *Foxc2* seem to function in the same systems as components of two signaling pathways converging on common genes, however, the compound heterozygotes do not present with all defects seen in each single homozygote, indicating that while these Foxc proteins function in the same systems, they do not have completely equivalent functions. This is best demonstrated by the fact that mutations in Foxc1 and Foxc2 cause different diseases.

Developmental pathogenesis

Msx2 and Alx4 expression is controlled by FOXC1 regulation of BMP signaling

To date, there is little known about *in vivo* targets of FOXC1. Recently, Foxc1 was suggested to regulate expression of homeobox gene *Msx2* and paired-typed gene *Alx4* by mediating *BMP* signaling (46). *Msx2* has a pivotal role in mediating the balance between early osteoprogenitor cell proliferation and differentiation in clavarial development. The expression pattern of *Foxc1* is consistent with roles during early and late stages of bone formation. In *ch* mice, the major skeletal defect is lack of clavarium associated with hydrocephalus. Rudimentary calvarial bones form at the sites of initial mesenchyme condensations but they fail to extend cranially (31, 43). *Msx2* -/- and *Alx4* -/- single homozygotes present with similar phenotypes, namely a delay in clavarial bone ossification and overall decrease in bone volume (46). *Ch* -/- mice exhibit a location-specific lack of *Msx2* and *Alx4* expression in the osteogenic clavarial mesenchyme and dura mater, suggesting a regulatory role for Foxc1. In *BMP2* induction assays *Msx2* was induced only in wild type and *ch* +/- mice, not in the Foxc1 null mutant calvarial explants, suggesting that while BMP

regulates *Msx2* and *Alx4* expression, this regulation is mediated by Foxc1 (46). It was also noted that Foxc1 is necessary for *Msx2* and *Alx4* expression only in proliferative calvarial mesenchyme (46).

Tbx1 is the first direct in vivo target of Foxc1

Tbx1 is a T-box transcription factor that may contribute to aortic arch development (47). TBX1 maps within the DiGeorge syndrome (MIM #188400) chromosomal region on 22q11.2. Tbx1 is expressed in the pharyngeal endoderm, pharyngeal mesoderm and broadly in the head mesenchyme (47). Foxa2, Foxc1, and Foxc2 all bind a cis-regulatory element in *Tbx1* and respond to *Shh* signaling to promote maintenance of Tbx1 expression (47). When this novel FOX binding site in the promoter region of Tbx1 is deleted, Tbx1 expression is abolished, resulting in aortic arch malformations. Foxc1-/- mice have similar aortic malformations and Tbx1 expression is reduced in these mice in the areas where Tbx1 and Foxc1 are co-expressed. This provides direct evidence for an *in vivo* target of Foxc1. In homozygous *Shh* null mutants, Foxc1 expression is down regulated followed by down regulation of Tbx1 expression.

Tbx1 expression is excluded from the neural crest-derived mesenchyme of the pharyngeal arches. Recent evidence suggests that Tbx1 may regulate fibroblast growth factor 8 (Fgf8) in the endoderm that can signal to adjacent neural crest cells and regulate aortic arch development (48).

Tyrosinase is a modifier to anterior chamber abnormalities

Mutations in the cytochrome P450 family 1, subfamily B, polypeptide 1 (CYP1B1) are a common cause of human primary congenital glaucoma. Studies of mutations in CYP1B1 led to the discovery of *Tyrosinase* (*Tyr*), a modifier of ocular drainage structure abnormalities (49). Tyrosinase converts tyrosine to L-hydroxyphenylalanine (L-dopa) and L-dopa to dopaquinone. L-dopa is the precursor to catecholamines, which are important developmental regulators. *Tyr-/- / Foxc1+/-* have severe ocular drainage defects compared to *foxc1* single heterozygotes. Administering L-dopa alleviated the severe angle dysgenesis present in double

homozygous *Tyr-/- / CYP1B1-/-* mice, suggesting a possible therapeutic role for tyrosinase.

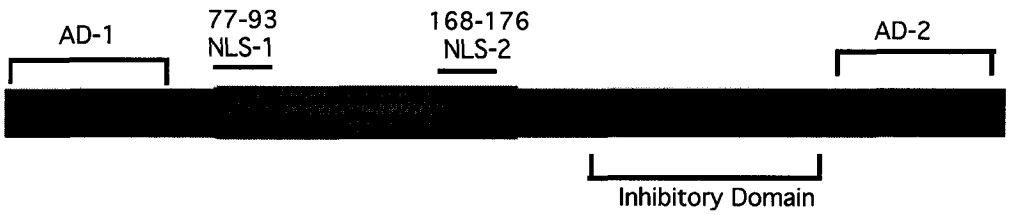
Foxc genes and the Notch signaling pathway

Kume *et al.* (35) propose that Foxc genes interact with the Notch signaling pathway to regulate the prepatterning of the anterior and posterior domains in the presumptive somites through a putative Notch/Delta/Mesp regulatory loop. It is thought that Mesp2 initiates segmentation in the presomitic mesoderm by down regulation of *Dll1* through the Notch signaling pathway (50). Embryos lacking either Foxc1 or Foxc2 or compound heterozygotes (Foxc1 +/- / Foxc2 +/-) die with similar defects in both the axial skeleton and cardiac systems, namely an absence of both early vascular remodeling and segmented paraxial mesoderm and somites (35). Embryos lacking pairs of closely related genes in the Notch pathway (Notch1-/- / Notch4-/- ; *presenilin1-/- / presenilin2-/-*) have defects in the cardiovascular system and somites similar to those found in Foxc1-/- / Foxc2-/- (35, 50), suggesting a role for Foxc genes in the murine Notch signaling pathway.

Molecular dissection of the FOXC1 FHD

Structure of the FOXC1 forkhead domain

The functional analysis of transcription factors has demonstrated that they are modular in structure, consisting of independently functioning protein domains (51) (52). *FOXC1* is under complex regulatory control as multiple domains are required for the correct targeting of FOXC1 to the nucleus and for efficient activation of a *FOXC1* transcriptional reporter gene (53). FOXC1 contains two regions required for nuclear localization (NL). The first region spans residues 77-93 but is not sufficient in itself to correctly target *FOXC1* to the nucleus; instead these residues act more as a nuclear localization accessory domain (53). The second region is necessary for NL and contains a stretch of residues at the C-terminal portion of the *FOXC1* FHD spanning residues 168-176 (53) (figure i-8). *FOXC1* also contains two transactivation domains (TD) at positions 1-55 and 466-553 (figure i-8).



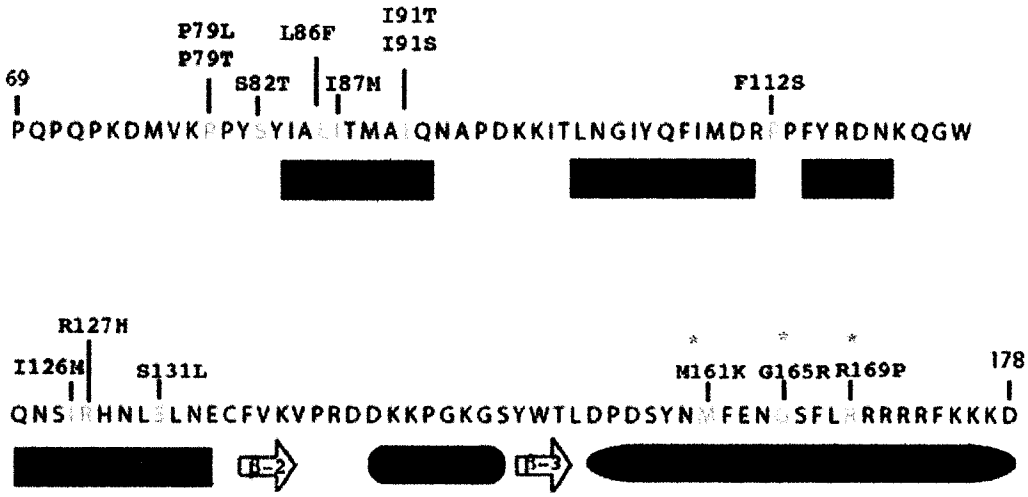
In addition to these transactivation domains, *FOXC1* harbors an inhibitory domain between the TDs at position 215-366. In the absence of this inhibitory region, *FOXC1* produces increased transactivation of a reporter gene construct (53).

Analysis of disease-causing missense mutations

To date, all disease-causing missense mutations identified in FOXC1 have been located in the FHD (figure i-9). Biochemical assays in conjunction with computational analyses were employed to determine the effects of novel patient mutations on FOXC1 function and structure (54) (55). Each mutation was introduced into the FOXC1 cDNA and the ability of the recombinant FOXC1 to localize to the nucleus, bind an *in vitro* derived FOXC1 binding site, and transactivate from a reporter construct was investigated. The mutations produce a spectrum of molecular defects. Analysis of these patient mutations has allowed assignment of putative functional roles for various regions of the FOXC1 FHD. Residues in α-helix 1 are involved in DNA binding, transactivation, and nuclear localization. The recognition helix H3, as determined by analysis of single mutations in this region, has no role in transactivation but has a major role in DNA-binding and nuclear localization (54, 55). These predictive assignments have allowed regions of the FOXC1 FHD to be considered in terms of function in conjunction with structure. The FHD is highly conserved and structural studies show limited three-dimensional variation between different FOX FHDs (56), therefore the functional assignments given to structures in the FOXC1 FHD are most likely to extend to other FOX proteins.

Evolution of wing 2

Winged helix/Fox proteins comprise a subset of the winged-helix-turn-helix (wHTH) proteins. wHTH proteins contain a classic HTH DNA-binding motif within a three helix bundle braced against an antiparallel β-sheet (57). In the winged helix/Fox protein subset of the wHTH family there is extension of a loop between β-strands that creates a wing-like structure that can be used to bind DNA (58). C-terminal to the β-sheet is another loop called wing 2, which is critical for Fox proteins to bind DNA as a monomer (57). The interactions between secondary elements



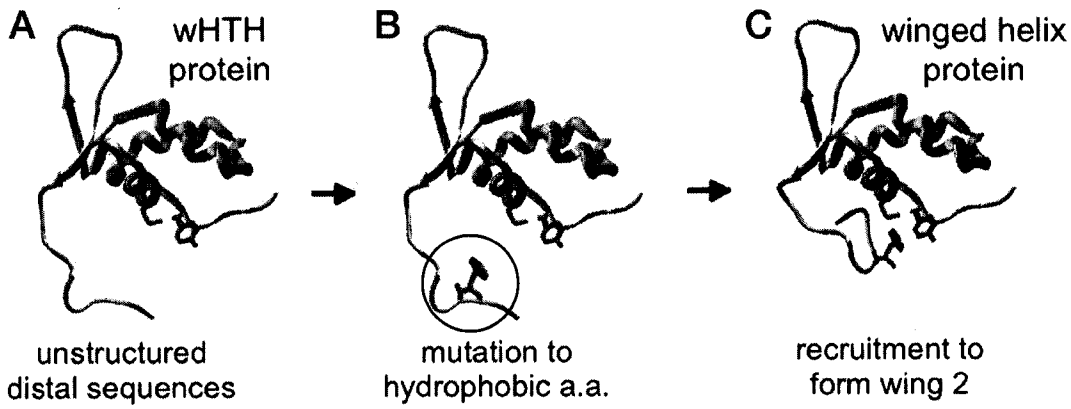
creates a three-dimensional structural entity that resembles a butterfly with the core comprised of the α-helices and β-sheets surrounded by two "wings", W1 and W2. This "winged helix" structure represents a structural unit which cannot be split further without losing its DNA binding abilities (25).

The presence of wing 2 is one of several features that distinguish Fox/winged helix factors from other wHTH proteins (57) and it is suggested that wing 2 and some amino acids upstream of this region evolved together to create the winged helix subdomain of wHTH proteins. Stabilization of wing 2 structure is essential for DNA binding (23, 57), and structural stability is highly sensitive to the presence of wing 2 hydrophobic and aromatic residues that are on the opposite face of the protein from that which contacts DNA (57). Stevens *et al.* propose a model for the evolution of the winged helix domain from a more general winged-helix-turn-helix structure (57). A primordial wHTH protein may have acquired, through mutation, bulky aromatic and/or hydrophobic residues in unstructured sequences that were distal to the wHTH domain. This distal polypeptide segment could then associate with corresponding aromatic and/or hydrophobic residues in the three-helix bundle that contribute to the stability or function of the wHTH domain (57). Additional mutations may enhance the association between segments of the domain and confer a functional advantage to the new protein (figure i-10). Some wHTH proteins function as dimers and Stevens *et al.* (57) suggest that the presence of wing 2 would extend the DNA binding face so that the protein may be able to bind the DNA effectively as a monomer.

DNA-binding specificity of other FHD family members

Sequence preceding the recognition helix H3 determines sequence specificity for FOXA3

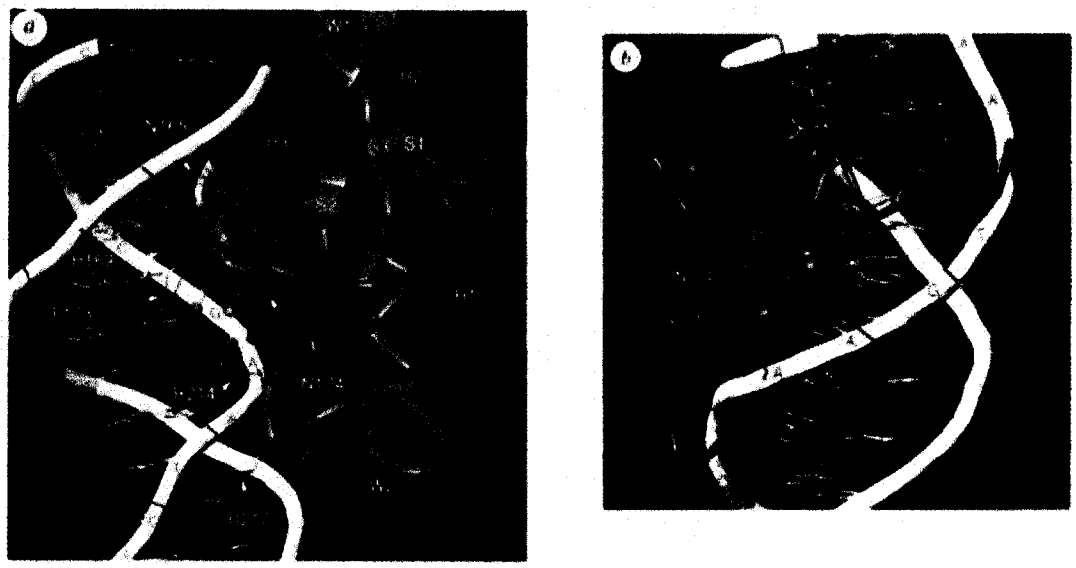
When the X-ray crystal structure of FOXA3 (formerly HNF3-γ) was determined (58), a binding oligonucleotide of thirteen residues (5' GACTAAGTCAACC 3') based on the FOXA3-DNA footprint and corresponding to positions –112 to -99 of the TTR promoter (60) was used in determining the

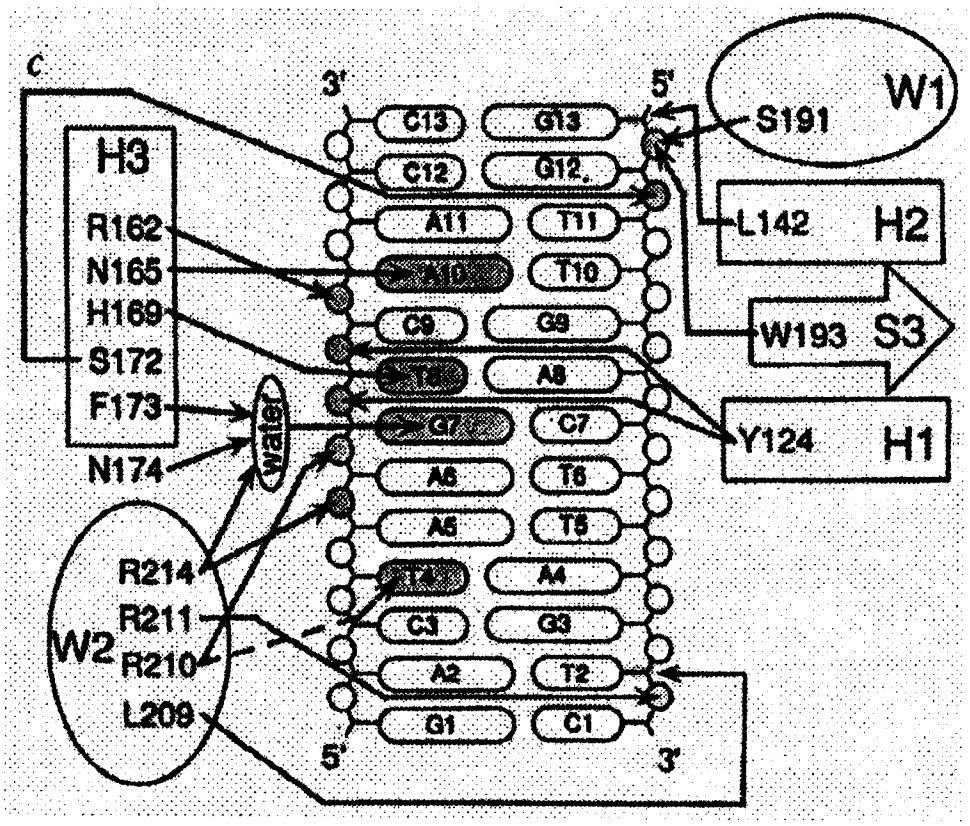


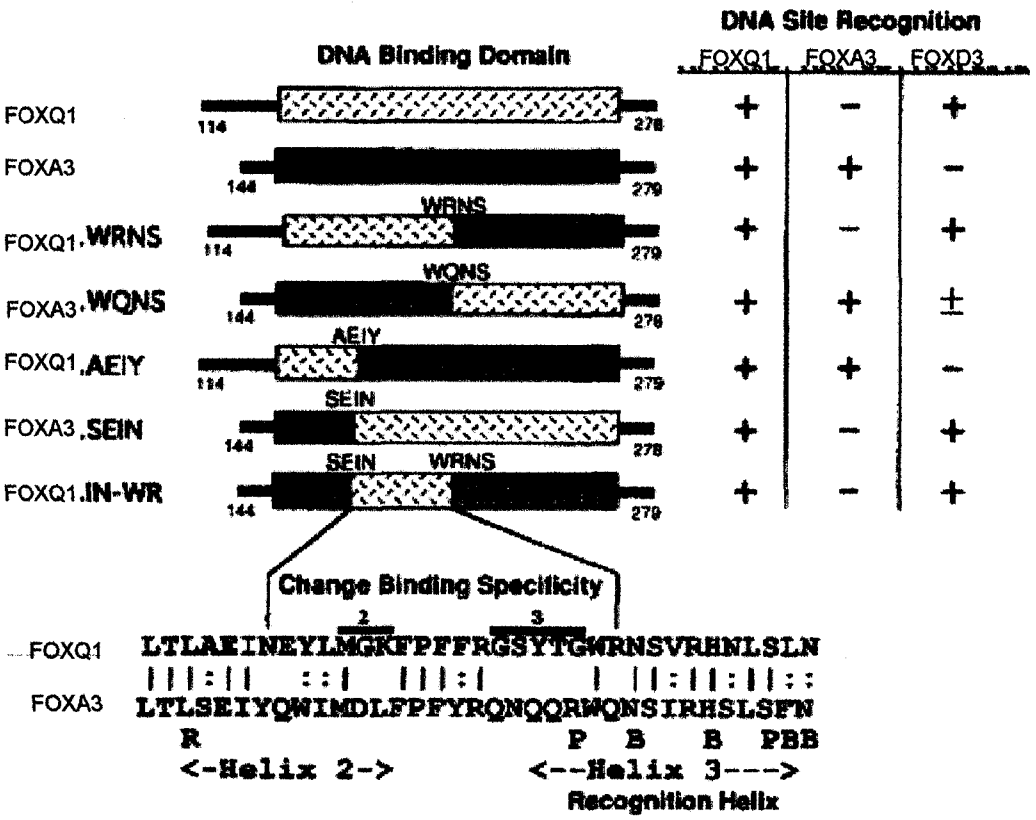
structure. FOXA3 binds DNA as a monomer and interacts with the DNA over a linear distance of 40 Å along the axis of the double helix, making contacts on both strands of the DNA (58). The recognition helix H3, lying in the major groove of DNA, provides the principal contact surface and makes the majority of direct base contacts as well as a number of phosphate backbone, ribose, and water-mediated contacts (figure i-11 a and b). Wing 2 makes several protein/DNA backbone and water-mediated contacts. Most importantly, wing 2 contacts nucleotide T4 in the minor groove from amino acid R210 (figure i-11 c). This is the sole direct base contact in the minor groove. FOXQ1, FOXD3 and FOXA3 have demonstrated distinct binding site recognition (61). The residues involved in base-specific contacts in helix 3 and wing 2 are conserved among family members and are therefore unlikely to directly contribute to conferring DNA-binding specificity. A series of domain swap constructs were designed to investigate amino acids outside these conserved regions that may play a role in determining binding specificity (figure i-12) (61). Binding assays that utilized the protein specific binding sites determined a segment of twenty amino acids that could alter sequence recognition abilities (figure i-12). This region encompassed residues from helix 2 to helix 3 and all intervening residues. Overdier *et al.* (61) demonstrated that this region of twenty amino acids, which is not homologous among family members, is capable of causing a recognition helix from one protein to bind DNA sites that are specific for another protein. From these data, a working hypothesis was formed where unique sequences preceding helix 3 were responsible for conferring sequence recognition properties to FOX proteins.

Creation of a small fourth α-helix H4, alters the position of the recognition helix, and thereby alters specific sequence recognition abilities of FOXD3

FOXD3 (formerly HFH-2 and Genesis) is a forkhead family member that has its structure determined both in a free state (56) and in a complex with DNA (62). The NMR structures of FOXD3 reveal some differences compared to the X-ray crystal structure of FOXA3 and suggest an alternative method of conferring DNA-binding specificity.





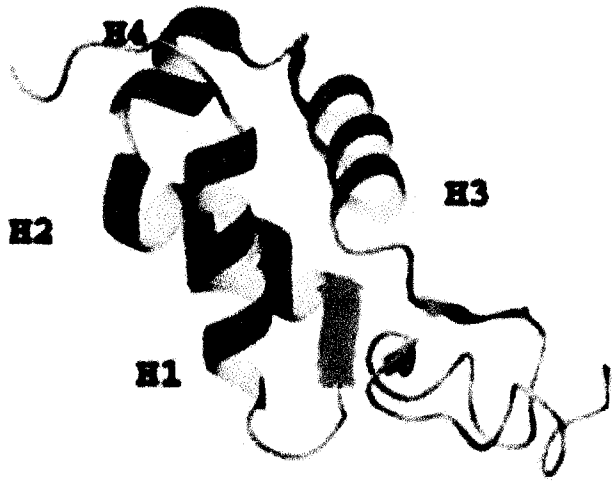


No significant changes are seen in the NMR structures of free *FOXD3* or *FOXD3* in a complex with DNA. The FOXD3 DNA-binding domain is 54% homologous to FOXA3, with the largest sequence variability residing in the seven residues preceding helix 3. These residues form a small fourth helix, H4. Several differences can be seen when the NMR structure of *FOXD3* is superimposed on the crystal structure of *FOXA3* (figure i-13). The largest difference in the superimposed structures is in the C terminal portion of the recognition helix, H3. This is noteworthy as this is the area where the majority of direct base contacts are made (56, 58). In the crystal structure of FOXA3 this stretch of twenty amino acids is twisted into a quasi-helical loop, and depending on the amino acid residues in this area, H4 may form (56). It is suggested that FOXD3 confers DNA-binding specificity by displacing the recognition helix, H3 through the formation of this helix 4 (56). In support of this, Marsden *et al.* (56) argue that this twenty amino acid region displays high sequence variability between FOX family members except for the middle section from Phe_{37}-Pro-Tyr-Tyr-Arg_{41} (numbering for FOXD3). This stretch of five amino acids is highly conserved in FOX family members and resides in the region where FOXD3 forms helix 4 and FOXA3 forms a quasi-helical loop. It is possible that these conserved residues have a propensity to form an α-helix and that the variable sequence flanking these residues determine the length of the helix, thereby varying interactions within this region of the FHD and subsequent interactions with the DNA (56).

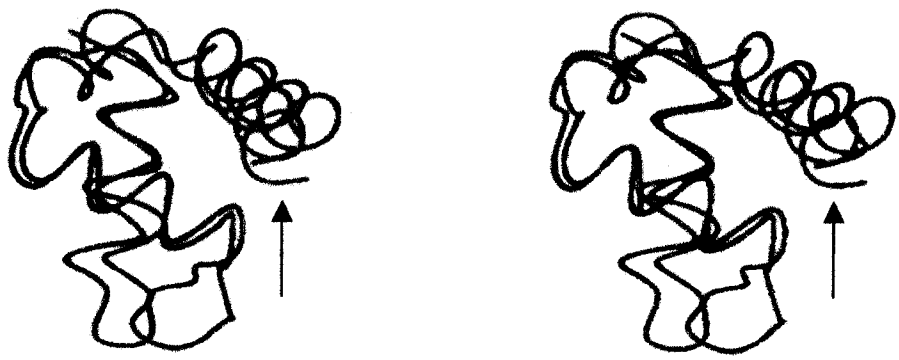
Analysis of binding specificity of several FOX proteins indicates important residues in FOXC1 and FOXD1 that control sequence recognition ability.

Pierrou *et al.* (63) investigated the possible relationship that sequence flanking the core binding site had in conferring DNA-binding specificity. Similar to domain swap experiments between FOXQ1 and FOXA3, a series of reciprocal swap constructs were designed containing various segments of FOXC1 and FOXD1 (figure i-14 a). The chimeras were tested against four different probes in a series of EMSA binding assays. Using a set of probes (figure i-14 b) that differed in the 3' sequence flanking the core consensus, Pierrou *et al.* found that sequence in wing 2 can influence the DNA-binding specificity of FOX proteins (63). When the swap

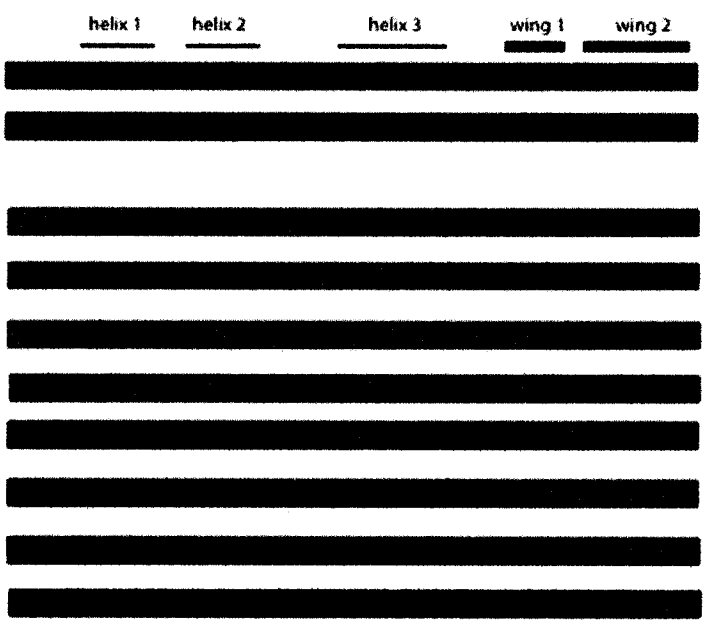
A



B



A



DNA SITE RECOGNITION

A	B	C	D
–	+	–	+
+	–	+	–
–	+	–	+
–	+	+	–
–	+	+	–
–	+	+	–
+	–	+	–
+	–	–	+
+	–	–	+
+	–	–	+

B

Probes

A	GATCCCTTAA**GTAAACA**GCATGAGATC
B	GATCCCTTAA**GTAAACA**AACAGAGATC
C	GATCCAGACT**GTAAACA**AAACATCGATC
D	GATCCAGACT**GTAAATA**AAACATAGATC

5'flank 3'flank

constructs were tested against a second set of probes that differed only in position +6 of the core consensus (figure i-14 b), specificity for this site in the core seemed to be determined from amino acids preceding helix 3. This is in direct agreement with the results of the domain swap experiments performed with FOXA3 and FOXQ1 (61), providing further support for the residues preceding helix 3 to be involved in determining binding specificity, as well as suggesting a role for the sequence in wing 2.

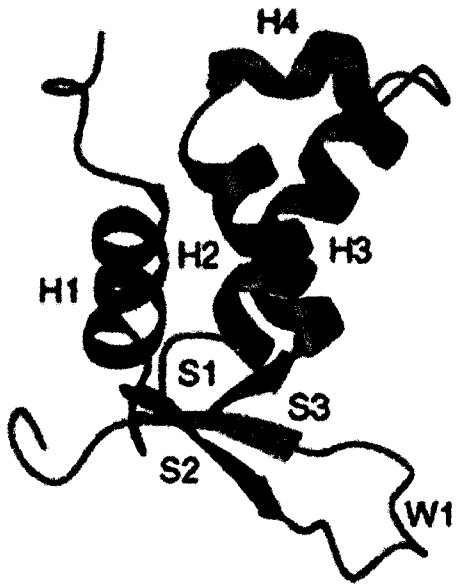
Differences in electrostatic potential mediate specific sequence recognition by FOXO4

The NMR structure for FOXO4 (formerly AFX1) was determined in 2001 and shows that FOXO4 adopts the expected winged-helix fold (64). Similar to FOXD3, the three helix bundle of FOXO4 is disrupted by a small, fourth helix, H4 (figure i-15 a). FOXA3, FOXD3, and FOXC2 share 65-70% sequence identity and are phylogenetically closely related, whereas FOXO4 shares only 40% sequence identity with this group and is more distantly related. Despite this low sequence identity, the structure of FOXO4 is similar to FOXA3 and FOXC2.

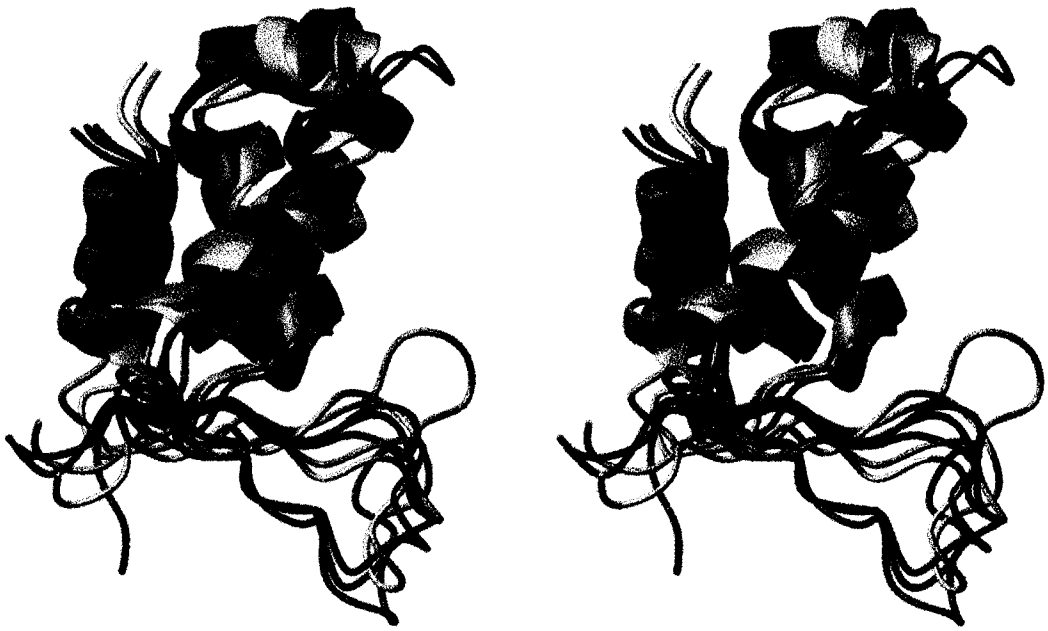
The presentation of the recognition helix to the DNA has been proposed as a way of conferring DNA-binding specificity (56). For FOXD3, creation of H4 may reposition the recognition helix thereby altering the presentation of residues to the major groove and subsequently altering sequence recognition. For FOXA3 and FOXD1, the mechanism of determining binding specificity has less speculation around structural orientation of helix 3 and more emphasis on sequence specific recognition based on indirect (non-base contacting) interactions from the variable region preceding the recognition helix. While FOXO4 has sequence variability in the region preceding the recognition helix, including the creation of the small fourth helix, the presentation of H3 is not structurally different from that of FOXA3 or FOXC2 (figure i-15 b). Therefore, based on previous hypotheses, there must be another factor involved in sequence recognition.

Weigelt *et al.* (56) suggest that while the sequence preceding helix 3 is important in determining binding specificity, it does so in a way other than through

A



B

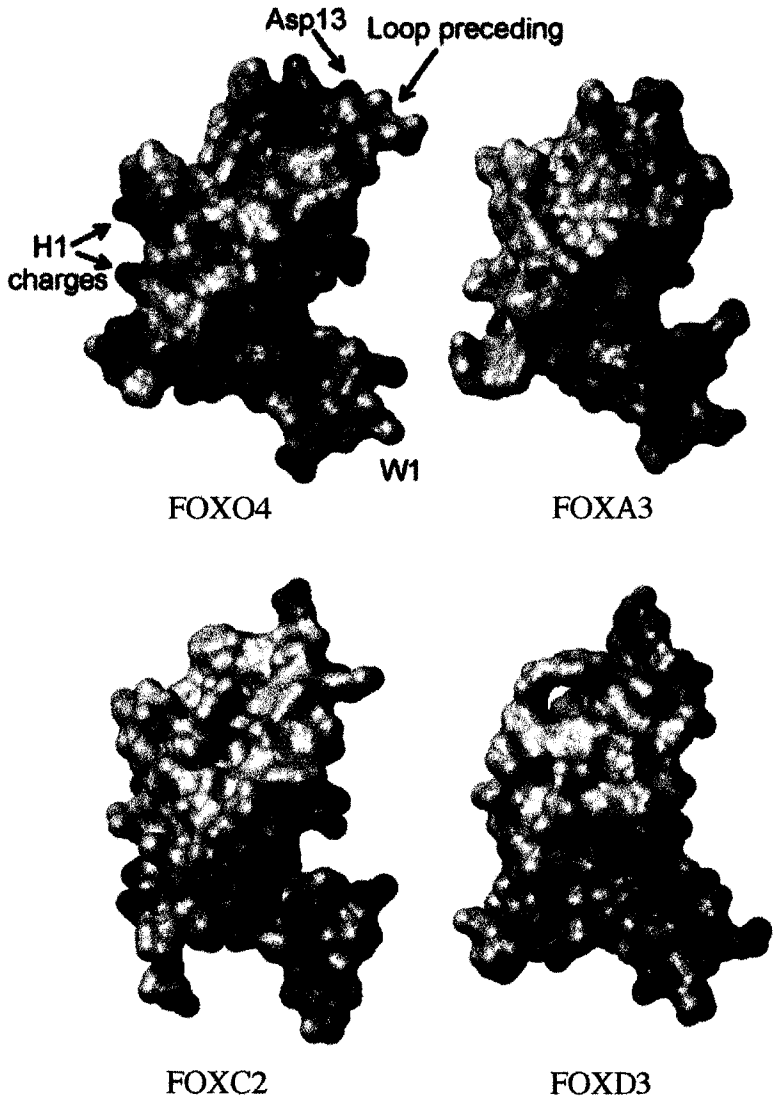


the formation of a small fourth helix (64). Instead, altered charge characteristics of the protein surface may affect the electrostatic interaction with the phosphate backbone of the DNA and thereby modulate the presentation of the recognition helix to the major groove (64). In comparing the electrostatic potentials of each protein face, it becomes obvious that there are differences in electrostatic potentials for FOXA3, FOXD3, FOXC2 and FOXO4 (figure i-16). Unique surface properties can be identified for each DNA-binding domain. FOXO4, for example, contains an additional negatively charged residue (Asp139) and a five amino acid insertion in the region preceding helix 3 (figure i-16). Two negatively charged residues in helix 1 H1, give rise to a negatively charged surface patch that is unique to FOXO4 (64). Instances where DNA binding is influenced by variations in charged residues have been reported (65). Therefore the molecular biology underlying DNA site recognition of the FOX class of transcription factors remains undefined.

Rationale for Project

FOXC1 is an important member of the FOX family of transcription factors, involved in numerous developmental systems. Mutations in *FOXC1* underlie AR malformations. Previous biochemical analyses of novel, disease-causing missense mutations have provided a functional role for some of the secondary structures of the FOXC1 FHD. However, one sub-domain that has not yet been functionally characterized is wing 2, a region for which the wHTH binding motif is named. As previously stated, wing 2 has evolved from a distal unstructured sequence to a sub-domain of structural significance (57). The identification of three novel disease-causing mutations in wing 2 provided an ideal system for me to investigate the functional significance of this poorly defined sub-domain and enhance the correlation between secondary structure and function in the FHD of FOX proteins.

While naturally occurring mutations provide a useful, biologically relevant system with which to investigate protein function, much more difficult is investigating the biochemical mechanisms used by proteins to control transcription



initiation, specifically in regards to recognition of specific DNA targets. The information I have presented in this general introduction indicates the intricacies involved in determining how a protein distinguishes one DNA target from another. Speculation surrounds the mechanism used by FOX proteins to discriminate between DNA sequences. With the exception of the rotated position of the recognition helix of FOXD3, the determined structures of FOX proteins in complex with DNA, show minimal structural differences (figure i-15). With few differences between the structures of FOX proteins, one has to wonder why the mechanisms suggested for recognizing target sequences are so different between FOX proteins.

Small patches of unique sequences have been implicated in conferring DNA recognition properties to FOX proteins (61). Subtle changes in electrostatic potential of the binding face have also been suggested as providing a means of specific DNA recognition (64). Repositioning of sub-domains, namely the recognition helix, which may alter indirect interactions with the DNA phosphate backbone is another suggestion for the mechanism by which FOX proteins may distinguish between sequences (56). FOX proteins are homologous and it is likely that the mechanism they use to recognize specific sequences is similar between family members. It is also probable that the mechanism used by FOX proteins to distinguish between sequences involves a combination of different recognition schemes. Further analysis into the mechanism of sequence recognition will most likely reveal that contributions from several of the recognition schemes suggested herein are involved in sequence discrimination.

In my analysis of FOXC1 function I was interested in gaining an understanding of the residues in *FOXC1* that contribute to the intricate task of sequence recognition. With the existing data of solved FOX protein structures, I investigated a region of twenty amino acids in the FOXC1 FHD that influences the sequence recognition properties of FOXC1. I used biochemical analyses in an attempt to determine the exact residues within this region that confer sequence recognition ability to *FOXC1*. It is likely that a task as complex as sequence discrimination is controlled by several factors and that my investigation into uncovering the residues involved in conferring recognition properties is in fact, one part of a whole system

necessary to recognize target sites. However, if the exact residues necessary for determining DNA-binding specificity can be determined, this may shed light onto the other requirements for target recognition.

Chapter One.

Analysis of Missense Mutations in Wing 2 of the FOXC1 Forkhead Domain

This chapter contains work published in:

The wing 2 region of the FOXC1 forkhead domain is necessary for normal DNA-binding and transactivation functions

Murphy TC, Saleem RA, Footz T, Ritch R, McGillivray B, and Walter MA. (2004) *IOVS* 45(8);2531-2538 (59)

Introduction

The importance of FOXC1 as a prominent transcription factor, with implications in numerous developmental processes, is without question. Its molecular function as a transcription factor, specifically in regard to the contribution of conserved residues to protein function, is being elucidated. The presence of naturally occurring, disease-causing missense mutations has provided a biologically relevant system with which to characterize FOXC1 function. Currently wing 2 is a poorly characterized region of the forkhead family of proteins. Therefore, I was specifically interested in analyzing novel mutations in this region as a way to gain dual insight into the molecular function of FOXC1 as well as the structural implications of wing 2.

Wing 2 is a distinguishing feature of Fox/winged helix-turn-helix factors and wing 2 sequence is essential for DNA binding (57). Structurally, wing 2 is in a position to interact with residues within the three helix bundle and studies suggest the importance of such intramolecular interactions to maintain the stability and function of wHTH proteins (57). I have previously stated the evolutionary significance of wing 2 and the selective advantages it may extend to the binding abilities of FOX/winged helix proteins. The presence of bulky hydrophobic residues in wing 2 reside in the vicinity of hydrophobic residues in the three helix bundle and actually expand the surface hydrophobic patch substantially (57). Clearly wing 2 has a fundamental role in DNA binding and protein stability.

In this chapter, biological assays were used in combination with computational homology modeling to understand the functional and structural significance of novel, disease-causing mutations in wing 2 of FOXC1. These analyses have shown wing 2 to be important for DNA-binding, where introduction of mutations caused a significant loss or reduction in binding ability. While it remains to be determined the extent wing 2 plays in determining DNA-binding specificity, I have demonstrated that single mutations in wing 2 are not enough to alter the binding specificity of FOXC1. In addition, I have demonstrated that wing 2 is important for transactivation, as introduction of mutations into this region cause a severe reduction

in ability of FOXC1 to drive expression from a reporter construct. FOXC1 is a known nuclear protein and I have demonstrated that single mutations that reside in the C-terminal NLS, which is necessary for proper nuclear localization, do not disrupt correct localization to the nucleus.

Methods

Plasmid construction and site-directed mutagenesis

The high G-C content of the FOXC1 cDNA did not allow for mutagenesis of the entire cDNA of FOXC1. For mutagenesis reactions a small fragment of FOXC1 cDNA, containing nucleotides 106-177, was PCR-amplified, using primers 5'-ggctacaccgccatgc- 3' (forward) and 5'- gctctcgatcttgggcac- 3' (reverse) and cloned into pGEM T Easy vector. FOXC1 was then mutated in pGEM T Easy using the Quickchange™ mutagenesis kit (Stratagene) and appropriate primers (Appendix A) following the manufacturer's protocol, with the addition of 5% DMSO. Mutagenesis products containing the correct mutation were selected and potential mutant constructs were sequenced using an automated sequencer (Li-COR). The fragment containing the correct mutation was then subcloned into pcDNA4A His/Max by way of an *ApaI – RsrII* digest and sequenced manually to confirm.

Cell transfection

Cells were grown in DMEM + 10% fetal bovine serum at 37°C with 5% CO_2. 100 mm plates of COS-7 cells were transfected at 80% confluence with 2.5 μg of plasmid using 20 μl FuGene6™ transfection reagent (Roche) as directed by the manufacturer. After 48 hours, proteins were harvested and analyzed by western analysis. For immunofluorescence COS-7 cells were grown on coverslips in 6-well plates and transfected with 1 μg of plasmid DNA with 3 μl of FuGene6. For transactivation assays, HeLa cells were grown in 6-well plates and transfected with 3 μl of FuGene6 using plasmid quantities listed below.

Protein extraction and western analysis

48 hours after transfection, cells were washed with PBS and harvested by scraping. Cells were pelleted at 2000 X g, resuspended in lysis buffer (20 mM HEPES pH 7.6, 150 mM NaCl, 0.5 mM DTT, 25% glycerol, 2.5 mM PMSF, 10 μg/μl aprotinin, 9μg/μl leupeptin, 10 μg/μl pepstatin A) at 4°C, and lysed by sonication on ice. Following centrifugation at 14 000 X g at 4°C for 5 minutes, supernatants were

transferred to a microfuge tube and resolved by SDS-PAGE. Resolved proteins were transferred onto nitrocellulose for western analysis. Mouse anti-Xpress™ antibody was used as the primary antibody for western analysis, with a goat anti-mouse conjugate to horseradish peroxidase as the secondary antibody.

Immunofluorescence

24 hours after transfection cells were washed three times with 1X PBS, followed by fixing for 10 minutes with 2% paraformaldehyde/PBS. Three washes with PBS were repeated followed by permeabilizing the cells using 0.05% Triton X-100/PBS for 10 minutes, followed by three washes in PBS. Cells were blocked for 1 hour using 500 µl per well of a 5% bovine serum albumin/PBS solution. After blocking, cells were washed three times with a 1% BSA/PBS solution followed by incubation for 1 hour with a 1:400 solution of the anti-Xpress antibody in the 1% BSA/PBS solution. Post incubation, the cells were washed three times with 1% BSA/PBS and then incubated with a 1:400 dilution of anti-mouse Cy3-conjugated secondary antibody (Jackson Immunolaboratories) in 1% BSA/PBS. Following an hour incubation, the cells were washed a final three times (1% BSA/PBS) and mounted with mounting media containing DAPI and the cover slips were sealed with clear nail polish.

Electrophoretic Mobility Shift Assays (EMSA)

Protein extracts harvested from COS-7 cells were quantitated by western blot analysis. Reactions were brought up to equal volume by the addition of untransfected COS-7 cell extract. Protein extracts were incubated with 1.3 mM DTT, 5 mg sheared salmon sperm DNA, 1 µg poly dIdC (Sigma), and 80 000-100 000 cpm of [^{32}P]-dCTP labeled double stranded DNA which contains the FOXC1 consensus binding site: forward 5'- gatccaaagtaaataaacaacaga -3'; reverse 5'- gatctctgttgtttatttactttg -3'). Reactions were incubated for 30 minutes at room temperature while the 6% polyacrylamide Tris-Glycine-EDTA gels used were pre-run for 15 minutes at 105V. Samples were then subjected to electrophoresis at room temperature for 50-60

minutes. Binding specificity EMSAs were performed as above with the addition of the [^{32}P]-dCTP labeled double stranded FOXC1 variant oligonucleotides.

Dual luciferase assay

The CMV promoter of pGL3 vector (Promega) was replaced with a Herpes Simplex Virus Thymidine Kinase (TK) promoter from the pRLTK (Promega) vector (54). The TK promoter was cloned into the *Bgl* II- *Hind* III sites of pGL3. Six copies of the FOXC1 binding site were then cloned into the *EcoR*I – *Nhe* I sites 5' to the TK promoter (54). HeLa cells were transfected with 50 ng of the pGL3 TK reporter construct, 1 ng of the Renilla control vector and 500 ng of a given FOXC1 pcDNA4A His/Max construct. Transfected cells were grown for 48 hours. The dual luciferase assays were performed using the Promega Dual Luciferase Assay kit according to the manufacturer's protocol (Promega).

DNA modeling

FOXC1 homology modeling was done using the SWISS-MODEL server (http://www.expasy.org/spdbv/ provided in the public domain by the Swiss Institute of Bioinformatics, Geneva, Switzerland).

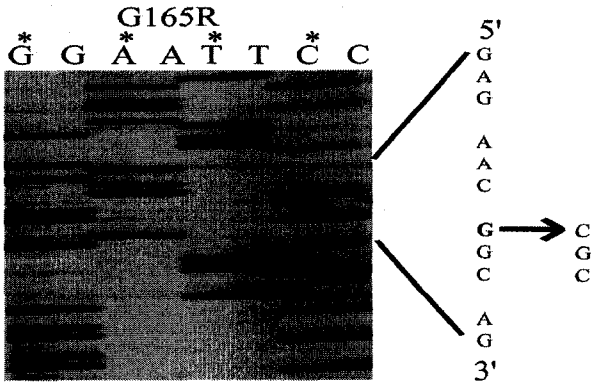
Results

Mutation detection in FOXC1

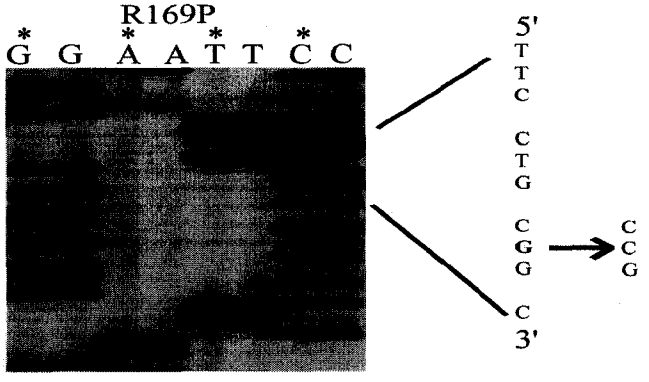
Mutation detection and sequencing of normal controls was a joint effort between Mrs. Farideh Mirzayans and Mrs. May Yu. Two single nucleotide changes were identified within wing 2 of the FHD of FOXC1 (figure 1-1) in two patients presenting with AR malformations. A G to C transversion at codon position 165 (494G>C; G165R) results in a glycine to arginine change. Another G to C transversion at codon position 169 (506G>C; R169P) results in an arginine to proline change. A previously reported M161K mutation, resulting from a T to A mutation at codon position 161 (482T>A; M161K) in two different AR patients (66) (67) was also included in the mutation analysis. Sequencing of the patients DNA and 100 normal control chromosomes confirmed that these mutations are not present in the normal population. Clinical photos were available for the patient harboring the R169P mutation (figure 1-2). The patient presented with iris hypoplasia, hypertelorism, corneal opacity, and abnormal pupillary function.

Mutagenesis and expression of FOXC1

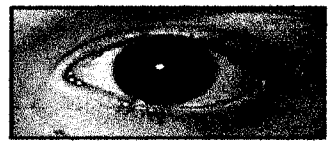
To correlate patient phenotype with FOXC1 function, the novel missense mutations in wing 2 were introduced into the FOXC1 cDNA and analyzed for disturbances to normal protein function. The missense mutations, G165R, M161K and R169P were introduced into the FOXC1 cDNA by PCR based site-directed mutagenesis on a small *ApaI – RsrII* fragment of the FHD. After the mutations were confirmed by sequence analysis, the mutagenized fragments were subcloned back into the FOXC1 pcDNA4 His/Max expression plasmid (Invitrogen). FOXC1 recombinant constructs were transfected into COS-7 cells and whole cell extracts were resolved by SDS-PAGE and western analysis. Protein was detected by use of an N-terminal (vector encoded) Xpress epitope that demonstrated a product of approximately 80 kDa in size (figure 1-3). Recombinant FOXC1 is approximately 65



Patient 1



Patient 2

R  L 

FOXC1 G165R M161K R169P

kDa; 61 kDa for the FOXC1 and 4 kDa for the Xpress tag. Wild type and mutant FOXC1 proteins were equalized by western blot band intensities.

Nuclear localization

As a DNA-binding protein, FOXC1 localizes to the nucleus (54). The signal necessary for FOXC1 nuclear localization (NLS) resides in the C-terminus spanning residues 168-176 (figure 1-4) (53). Of the three missense mutations investigated, R169P resides within the C-terminal NLS while the remaining two mutations (M161K and G165 R) reside at the periphery. The effect of these wing 2 missense mutations on proper FOXC1 localization was tested in COS-7 cells transiently transfected with the FOXC1 pcDNA4 His/Max vector missense mutation constructs.

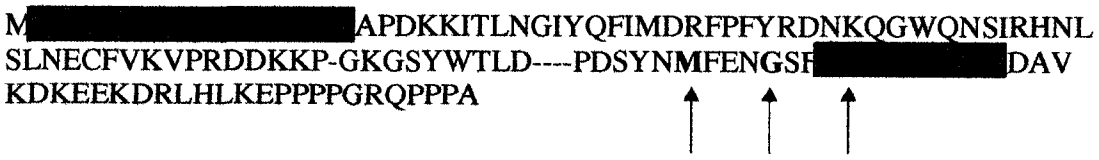
Immunofluorescence detection of Xpress tagged (Stratagene) recombinant FOXC1 showed that wild type FOXC1, G165R, M161K and R169P all localize predominantly to the nuclei of COS-7 cells with 96%, 91%, 88% and 88% nuclear localization, respectively (figure 1-5). Introduction of these mutations into FOXC1 cDNA are not sufficient to disrupt proper nuclear localization.

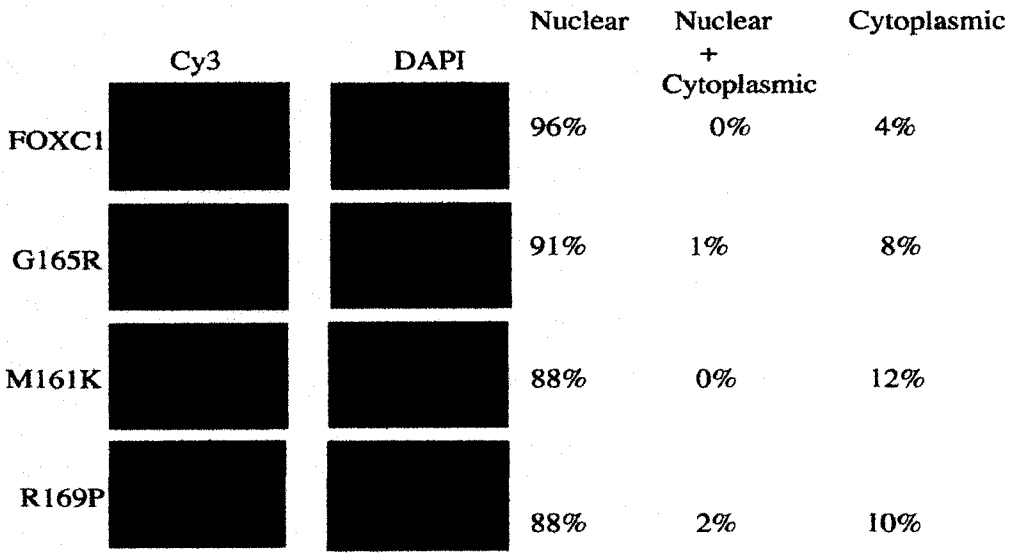
Analysis of wing 2 DNA-binding capacity and specificity

Electrophoretic mobility shift assays (EMSAs)

EMSAs indicate that FOXC1 forms a preferred DNA-protein complex with an *in vitro* derived oligonucleotide (63) (Table 1-1). This FOXC1 binding site was utilized to determine how recombinant FOXC1 binding was affected by the introduction of missense mutations in wing 2.

The G165R mutation showed a capacity for DNA binding comparable to wild type FOXC1, while M161K and R169P mutations showed a reduction in DNA-binding by 2-fold and 5-fold of wild type levels, respectively (figure 1-6).





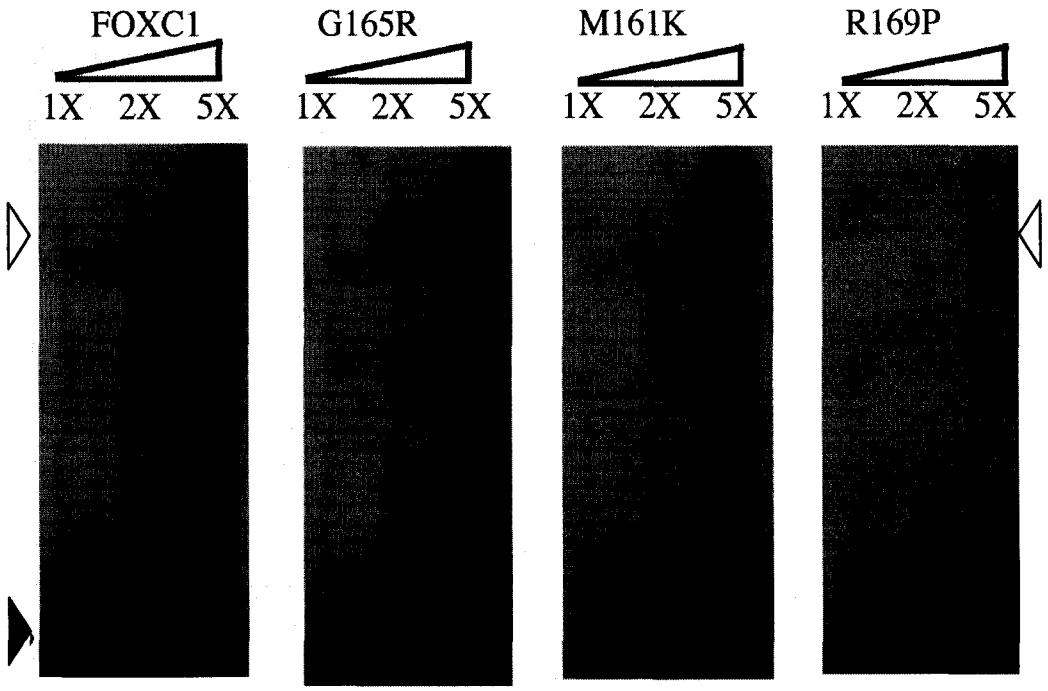


Table 1-1. The FOXC1 binding site and variant oligonucleotides

Oligonucleotide	Forward	Reverse
FOXC1 Binding site	gatccaaagtaaataaacaacaga	gatctctgttgtttatttactttg
1	gatccaaactaaataaacaacaga	gatctctgttgtttatttagtttg
2	gatccaaagaaaataaacaacaga	gatctctgttgtttatttttctttg
3	gatccaaagataataaacaacaga	gatctctgttgtttattaactttg
4	gatccaaagtatataaacaacaga	gatctctgttgtttatatactttg
5	gatccaaagtaattaaacaacaga	gatctctgttgtttaattactttg
6	gatccaaagtaaaaaaacaacaga	gatctctgttgttttttactttg
7	gatccaaagtaaattaacaacaga	gatctctgttgttaatttactttg
8	gatccaaagtaaatatacaacaga	gatctctgttgtatatttactttg
9	gatccaaagtaaataatcaacaga	gatctctgttgattatttactttg

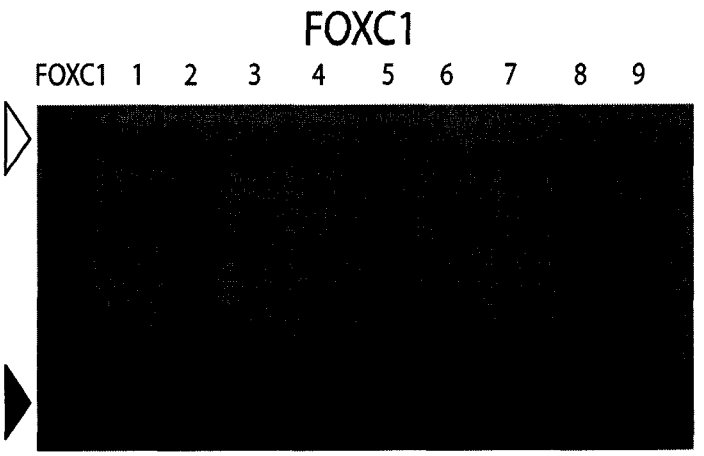
DNA-binding specificity assays

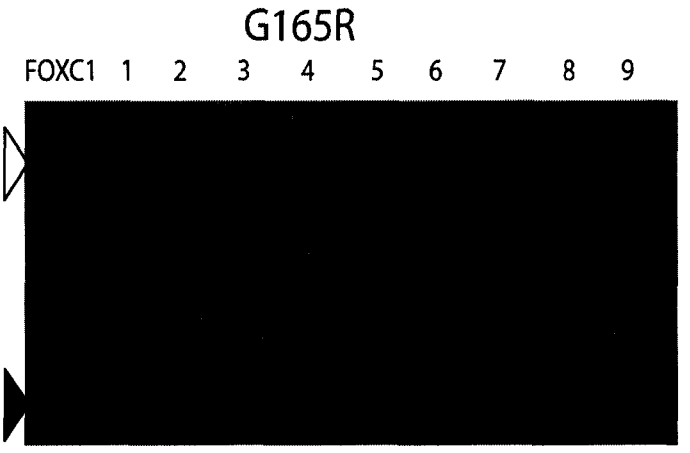
FOXC1 is known to bind a series of variant FOXC1 oligonucleotides with a specific affinity for each site (54). Wing 2 missense mutations, G165R and M161K, maintain binding capacity for the consensus FOXC1 binding site. Therefore to determine if the introduction of missense mutations in wing 2 would alter the DNA-binding specificity of FOXC1, these recombinant FOXC1 constructs were tested against the panel of variant FOXC1 binding sites (Table 1-1). The affinity of each recombinant protein for each altered binding site was compared to the wild type affinity for the variant sites. The reduction in binding ability of the R169P mutation (5-fold reduction in affinity) most likely results from the loss of contact with the minor groove upon introduction of a proline at this site. As DNA contacts are required to maintain binding ability, and R169P has lost this ability, R169P was not included in the binding assays against the variant oligonucleotides.

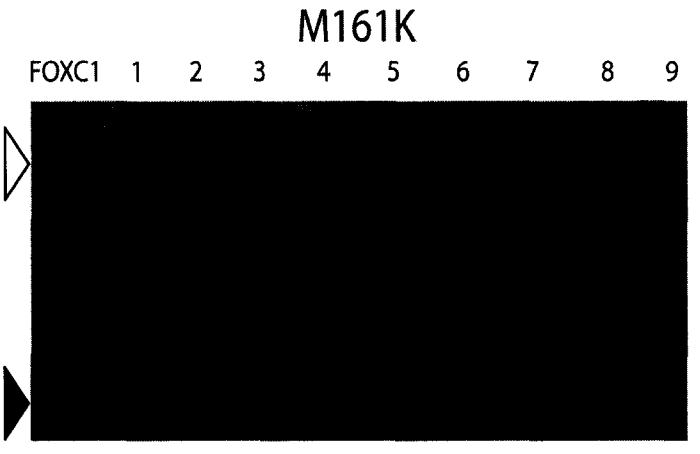
Both G165R and M161K show a pattern of affinity for the variant oligonucleotides equivalent to that of wild type FOXC1 (figure 1-7) and therefore do not alter FOXC1 sequence preferences to this panel.

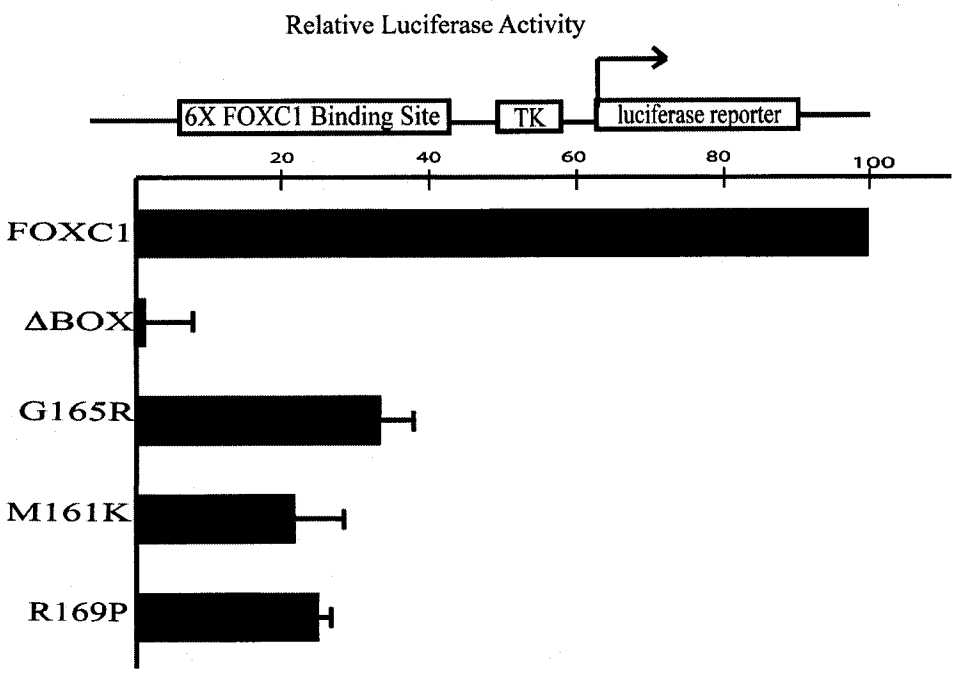
Transactivation assays

FOXC1 has been shown to transactivate gene expression of a reporter construct (53) (54) thus the effect of missense mutations G165R, M161K, and R169P upon the transactivation potential of FOXC1 was investigated. A reporter construct with six FOXC1 binding sites positioned upstream of a Herpes Simplex Virus thymidine kinase (TK) promoter was used to determine if the recombinant FOXC1 constructs were able to activate expression from the luciferase reporter. It has been previously established that the addition of the six FOXC1 binding sites upstream of the TK promoter results in 12.5 fold greater activation of the luciferase reporter by FOXC1 compared to the TK promoter alone (personal communication, Dr. R. Saleem). HeLa cells were co-transfected with the TK-luciferase reporter construct and pcDNA4 His/Max or FOXC1 pcDNA4 His/Max. FOXC1 in which the FHD was deleted (ΔBOX) activated luciferase expression at only 1% of wild type levels (figure 1-8). The ability to activate the luciferase gene was disrupted by all three missense





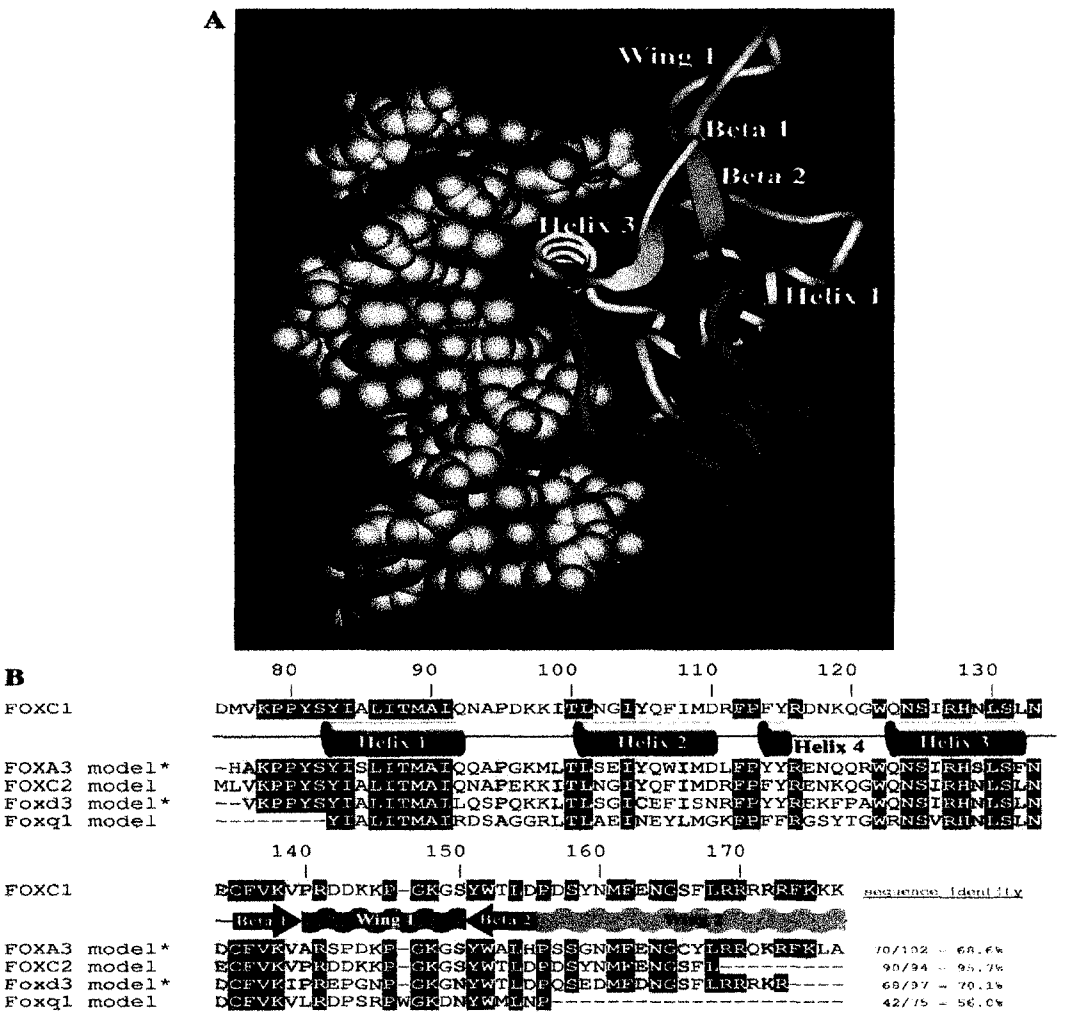


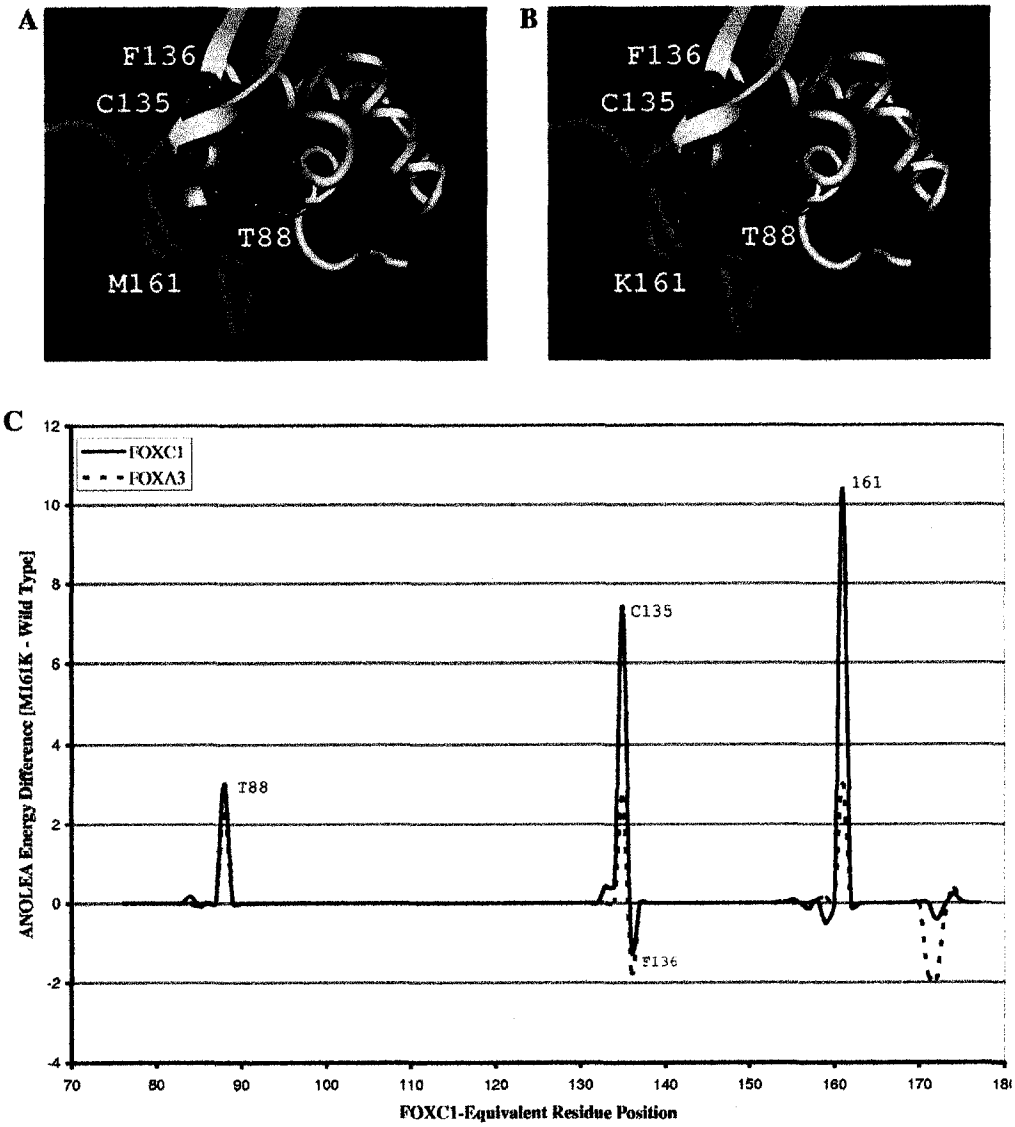


mutations (figure 1-8). FOXC1 carrying G165R transactivated expression of the luciferase reporter at 36% of wild type levels. M161K and R169P levels of transactivation were 20% and 23% of wild type, respectively. These results indicate that although these mutant proteins are able to bind to the FOXC1 binding site *in vitro* and maintain nuclear localization, all three mutant proteins exhibit reduced levels of gene activation. Therefore the overriding defect resulting from all three mutations is likely to be an impaired capability to activate transcription of target genes.

Molecular modeling

Molecular modeling and analysis was done by Tim Footz. A theoretical structure of the forkhead domain of FOXC1 was generated by homology modeling using FOXA3 (figure 1-9A) as the template. Of the closely-related forkhead domains with known structure, the FOXA3 model has the most complete wing 2 region and also contains identical residues at the positions equivalent to M161, G165 and R169 in FOXC1 (figure 1-9B). The model structures were submitted to the ANOLEA (Atomic Non-Local Environment Assessment) (68) server which computes a non-local energy profile for assessing protein structure. Methionine 161 in FOXC1 was mutated to lysine *in silico* to predict structural defects in the M161K molecule. Figure 1-10 illustrates that the M161K mutation may disrupt structurally-favorable interactions in an interior region of FOXC1 where T88, C135, F136 and M161 are predicted to converge. Assuming no changes occur to the structure of the protein backbone upon mutagenesis, the optimal rotamer for K161's side chain noticeably increases the ANOLEA scores of residues 88, 135 and 161 (figure 1-10C). This pattern suggests that these residues are normally involved in a highly-organized region of side chain packing (figure 1-10A) such as would occur in a globular protein's hydrophobic core. Disruption of this core-like module to accommodate the M161K amino acid substitution may lead to localized instability or deformation of wing 2. *In silico* mutagenesis to create G165R and R169P models did not result in any significant changes to ANOLEA scores of non-local residues (not shown) as expected for solvent-exposed positions.





Discussion

Missense mutations in wing 2 disrupt FOXC1 function

Mutations in wing 2 of the FOXC1 FHD, or in any FOX proteins, have not been previously subject to molecular analyses. A comprehensive summary of the molecular defects caused by missense mutations in FOXC1 is listed in Table 1-2. AR patients with mutations in wing 2 of FOXC1 present with phenotypes similar to those seen in AR patients with mutations in other regions of the FHD (54, 55). Thus, there appears to be no phenotype/genotype correlation between the region of mutation in the FHD and patient phenotype.

Nuclear localization is not disrupted by mutations G165R, M161K or R169P

The nuclear localization signals (NLS) of FOXC1 reside in the N and C-terminus of the FHD, encompassing residues 77-93 and 168-176, respectively (53) (figure 1-4). M161 and G165 reside outside of the C-terminal NLS, therefore their lack of disturbance on protein localization is not surprising. R169 does reside in the C-terminal region required for nuclear localization, however, substitution at this site to P169 is not sufficient to disrupt the signal and perturb nuclear localization. Therefore the above wing 2 mutations are tolerated by the C-terminal NLS and do not disrupt FOXC1 localization to the nucleus.

G165R maintains wild type levels of DNA binding but disrupts transactivation ability

In the FOXA3/DNA crystal structure position G165 extends away from the binding interface and is not a predicted base, phosphate backbone, or water mediated contact. (58) Therefore, mutation at position 165 is not predicted to disrupt DNA binding. The fact that FOXC1 carrying G165R can maintain wild type levels of DNA binding (figure 1-6) agrees with this prediction and indicates that the winged-helix structure necessary for DNA binding is not disrupted. The severe reduction in transactivation due to the G165R mutation however, implies a functional role for this position in transactivation. Previous characterization of two FOXC1 disease-causing missense mutations, F112S and I126M in the FHD of FOXC1 show similar

Table 1-2. Summary of molecular defects caused by missense mutations in FOXC1

	Molecular Defect				
Mutation & 2 structure Involvement	Nuclear	DNA binding capacity	TA	Altered binding specificity	Reference
P79L N-terminal Region	++	+	+	NO	Saleem *et al.* 2003
P79T N-terminal Region	++	+	+/-	NO	Saleem *et al.* 2003
S82T N-terminal Region	+++	+	+	NO	Saleem *et al.* 2001
L86F α-helix 1	+++	++	-	NO	Saleem *et al.* 2003
I87M α-helix 1	Reduces Protein Stability				Saleem *et al.* 2001
I91S α-helix 1	-	+	-	NO	Saleem *et al.* 2003
I91T α-helix 1	+	+	-	NO	Saleem *et al.* 2003
F112S T-loop	+++	+++	-	NO	Saleem *et al.* 2001
I126M α-helix 3	+++	+++	-	YES	Saleem *et al.* 2001
R127H α-helix	-	-	-	NO	Saleem *et al.* 2003
S131L α-helix 3	++	+	-	NO	Saleem *et al.* 2001
M161K Wing 2	+++	++	+/-	NO	Murphy *et al.* 2004
G165R Wing 2	+++	+++	+/-	NO	Murphy *et al.* 2004
R169P Wing 2	+++	+	+/	N/A	Murphy *et al.* 2004

Listed above are missense mutations that have been studied to date and the molecular effects of the mutation. TA: transactivation. Nuclear localization and transactivation scoring: 81-100% (+++), 61-80-% (++), 41-60% (+), 21-40% (+/-), 0-20% (-). DNA binding scoring: 1X (+++), 2X-4X (++), 5X-9X (+), 10X (+/-), <10X (-). binding (54).

disruptions in transactivation while maintaining normal levels of DNA binding (52). Disruptions of protein/protein interactions by the G165R mutation may disrupt intermolecular interactions necessary for transcription activation. Previous reports suggest mutations F112S and I126M do not disrupt DNA binding because they are not involved in protein/DNA interactions (54). Rather, they disrupt transactivation possibly due to disturbances in protein/protein interactions and/or intramolecular interactions (54). Previous characterization of the FOXC1 FHD has implicated the N-terminus of α-helix 1 in transactivation and DNA binding (54). Wing 2 of FOXA3 meanders across the surface of the three helix bundle in the vicinity of the N terminus of α-helix 1 establishing an environment receptive to interaction between regions of α–helix 1 and wing 2 (58). Potential intramolecular interactions between wing 2 and α-helix 1 may be disrupted by the G165R mutation, maintaining DNA-binding ability while indirectly disrupting the transactivation ability.

M161K has a minor reduction in DNA binding ability coupled to a large disruption of transactivation ability

Stable formation of wing 2 is required for DNA binding of FOX proteins (58). Figure 1-10 illustrates that the M161K mutation may disrupt structurally-favorable interactions in an interior region of FOXC1 where T88, C135, F136 and M161 are predicted to converge. In our predicted structure of FOXC1, M161 is spatially arranged such that its side chain resides in close proximity to the side chains of T88, C135 and F136 in a region of reduced solvent accessibility (figure 1-10), where the cusp of wing 1 converges with α–helix 1 and the middle of wing 2. The substitution of polar lysine for nonpolar methionine at position 161, as occurs in both patients with the reported mutation, places a larger positively charged residue in the vicinity of a hydrophobic core-like module. *In silico* mutagenesis of M161 to K161, in which only the atomic constituents and rotation of the amino acid side chain at position 161 were altered, suggests that the new side chain would no longer closely associate with the T88/C135/F136 cluster (figure 1-10). M161K may perturb the stability of wing 2 enough to reduce the DNA-binding ability by affecting the interatomic non-local interactions between residues T88, C135 and M161.

Unlike G165R, the loss of transactivation in M161K is most likely due to the approximate 2-fold reduction in DNA binding of this mutation (figure 1-6). M161K is similar to a previous mutation, L86F, in α-helix 1 of the FHD of FOXC1 (69). L86F shows a minor reduction of DNA binding ability with a similar reduction in transactivation from a luciferase reporter, implicating L86F as having an important role in intramolecular interactions (69). In FOXA3, position M202 is one of thirty-five known residues that contribute to the hydrophobic core of the FHD (58). This position is conserved in FOXC1 and corresponds to position M161. Therefore M161 may have a role in intramolecular interactions that the M161K mutation disrupts, impairing transactivation.

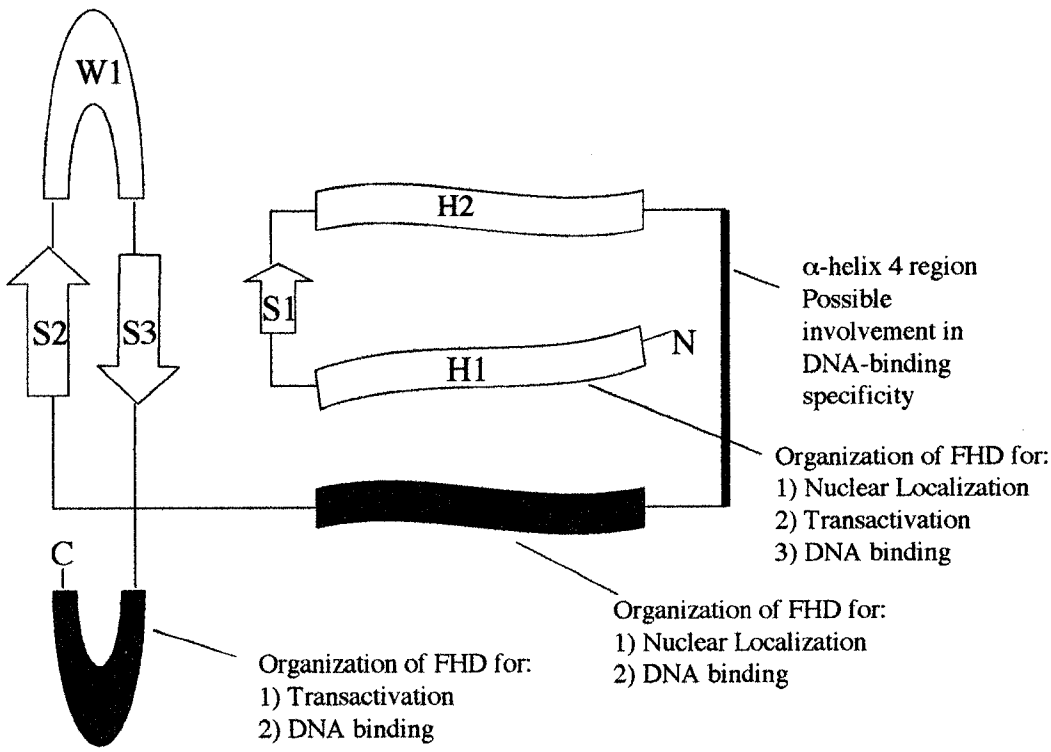
Analysis of position R169, the sole forkhead domain side chain-base contact in the minor groove.

Based on the crystal structure of FOXA3 (58), the R169P mutation occurs in a position that is predicted to be responsible for the sole side chain-base minor groove contact in wing 2 (58). R169P shows a severe loss of binding capacity, over a 5-fold reduction in affinity for the FOXC1 binding site (figure 1-6). My analysis predicts that mutation at position 169 disrupts the direct base contact, potentially destabilizing the formation of wing 2 which is critical for DNA binding activity (58).

The severe reduction in DNA binding from mutation R169P is translated into a marked reduction in transactivation of a luciferase reporter (figure 1-8). The severe reduction in DNA binding seen in FOXC1 R169P, and the knowledge that R169 makes a DNA contact in the related FOX protein FOXA3 (58), supports the hypothesis that this reduced transactivation is a direct result of reduced DNA binding capability.

Analysis of wing 2 as a functional sub-domain of the FOXC1 FHD

Previous analyses of the FOXC1 FHD have concentrated on novel, missense mutations that reside in α-helix 1, α-helix 3 and the T-loop between α-helices 2 and 3 (54, 55, 69). These analyses have allowed putative functional roles to be assigned to the sub-domains of the FOXC1 FHD (figure 1-11). The recognition helix is



important in DNA-binding, however in-depth analyses also indicate that helix 3 is necessary for proper localization to the nucleus. α-helix 1 is suggested to play a role in transactivation, as well as binding. Functional assignment has allowed a more detailed structure/function relationship to develop for the sub-domains of the FHD. With further modeling and biochemical analysis, the structure of the FHD and the functional roles of the sub-domains will become refined. Previous work indicated wing 2 was necessary for maintaining protein stability and suggested a possible role in DNA binding (57) (63). My analysis of the three missense mutations, G165R, M161K, and R169P, has helped to elucidate the functional significance of the wing 2 sub-domain. My biochemical analyses have shown that residues in wing 2 that make direct (R169P) and indirect (M161K) interactions with the DNA are important to maintain proper levels of DNA binding. The loss of binding upon mutation of position R169 provides direct evidence for wing 2 having a functional role in the binding capability of FOX proteins. In addition to the role wing 2 plays in DNA binding, my results also suggest a functional role for wing 2 in transcription initiation, namely activation of target genes. The reduction of transactivation ability with the G165R mutation, a mutation that does not disrupt binding ability, suggests that, aside from making necessary contacts with the DNA, wing 2 is involved in inter/intramolecular interactions necessary for proper gene activation. Clearly, this is a region that has structural as well as functional significance in the family of wHTH proteins.

Chapter Two.

Analysis of DNA-binding specificity of FOXC1

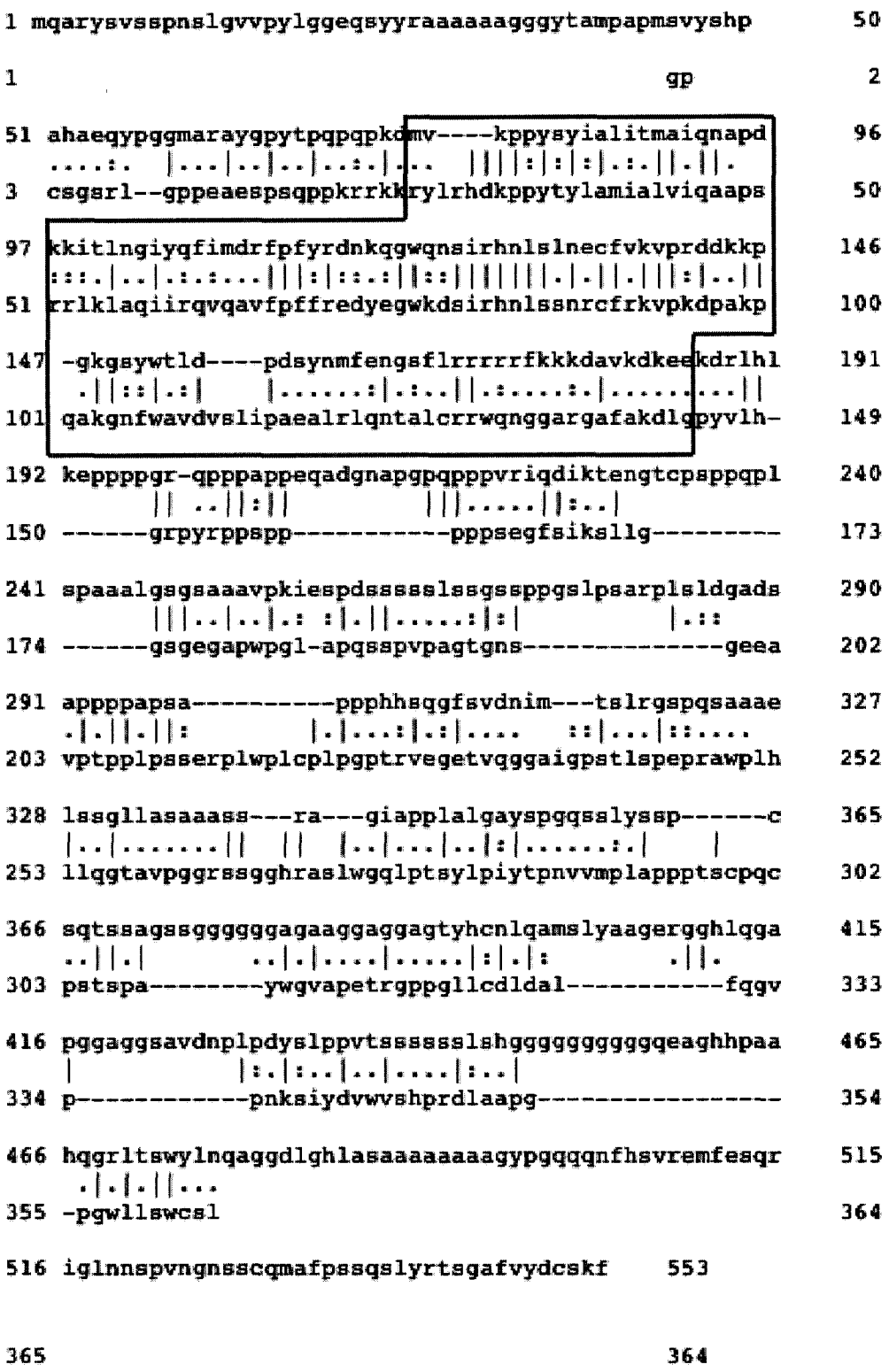
Introduction

Several control mechanisms operate at the level of transcriptional initiation, however one of the most critical steps is the recognition of a specific DNA sequence by transcription factors, leading to enhanced and/or repressed transcription of target genes (70). This recognition process requires both base-specific and non-specific interactions between amino acids of the transcription factors and the nucleotide residues of their target sequences (70). The FOX/winged helix binding domain is highly conserved within this family of transcription factors, yet each protein has distinct DNA binding abilities. To fully understand the function of FOX proteins it is essential to comprehend the DNA contact schemes employed by FOX family members. In particular, it is important to understand whether or not homologous FOX members use the same mechanisms to control sequence recognition or whether there are differences among family members.

As mentioned previously, several forkhead family members have had their structure determined by NMR or x-ray crystallography. From these structures several mechanisms for determining binding specificity have been suggested (70) (56) (63) (57). To complement my work on the molecular investigation of the function of the FOXC1 FHD through analysis of wing 2, I initiated a parallel line of study into elucidating the residues FOXC1 employs to determine DNA-binding specificity.

A common approach to analyze the specifics of DNA-binding is through domain swapping experiments (61) (63). Switching the controlling sequences of proteins can alter their ability to discriminate against DNA sequences. To study the DNA-binding specificity of FOXC1, I built swap constructs between FOXC1 and a distantly related forkhead protein, FOXH1.

FOXH1 was selected as the reciprocal protein in my FOXC1 domain swap experiments because it is a distantly related forkhead protein with only 20% identity in the central region of the FHD (figure 2-1). Unlike FOXC1, which can transactivate expression from a reporter construct by co-transfection of the expression construct



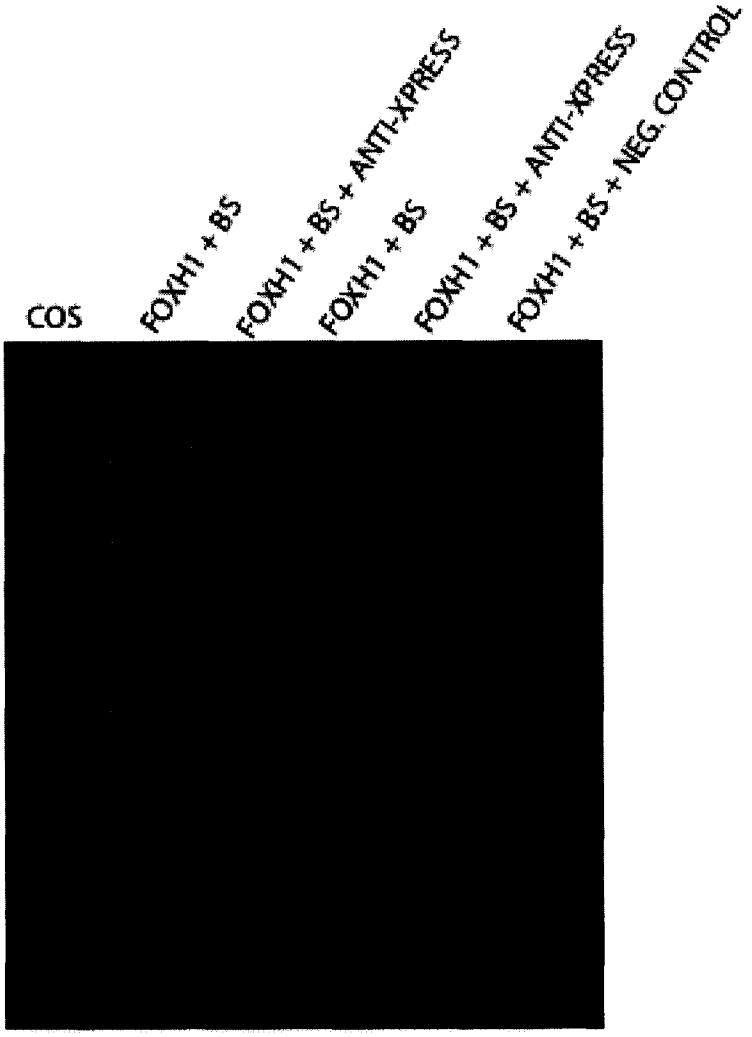
with the reporter construct alone, FOXH1 requires additional elements, such as TGF-β and SMAD 4 (71). Site selection and amplification studies indicate that FOXH1 binds an 8 bp consensus binding site TTA(T/G)(T/G)TGT (71). This sequence differs slightly from the general suggested core consensus for forkhead proteins, however the ATAA core is present (figure 2-2b). As a positive control for domain swapping experiments, FOXH1 was suboptimal as binding of FOXH1 to its *in vitro* derived binding site was weak at best (figure 2-2a). As a result, in my domain swap experiments, FOXH1 sequence was introduced into FOXC1 cDNA but the reciprocal changes were not introduced into FOXH1.

Previous studies implicate a region preceding helix 3 to be involved in determining DNA-binding specificity of forkhead proteins (61), therefore a region of twenty-one amino acids that precede the recognition helix, H3, of FOXC1 were swapped in varying segments to create a total of six swap constructs (figure 2-3). Due to the significant amount of difference in the FOXC1 and FOXH1 sequence in this region (fourteen amino acids differ out of the total twenty-one) the mutagenesis was divided into three separate segments, each concentrating on a region of seven consecutive amino acids. This resulted in three constructs containing a different set of seven swapped amino acids, two constructs containing a different segment of fourteen swapped amino acids, and a final construct containing all twenty-one amino acids swapped. (Swap construct C was built by Dr. Ramsey Saleem).

In an attempt to discern the least number of amino acids required to alter FOXC1 binding, based on analyses of the constructs described above, four additional swap constructs were created that harbored a different, single amino acid change in the central region of the twenty-one amino acids initially under investigation (figure 2-4). A swap construct that replaced the entire recognition helix H3, of FOXC1 with that of FOXH1 was created in the same manner as those describe above.

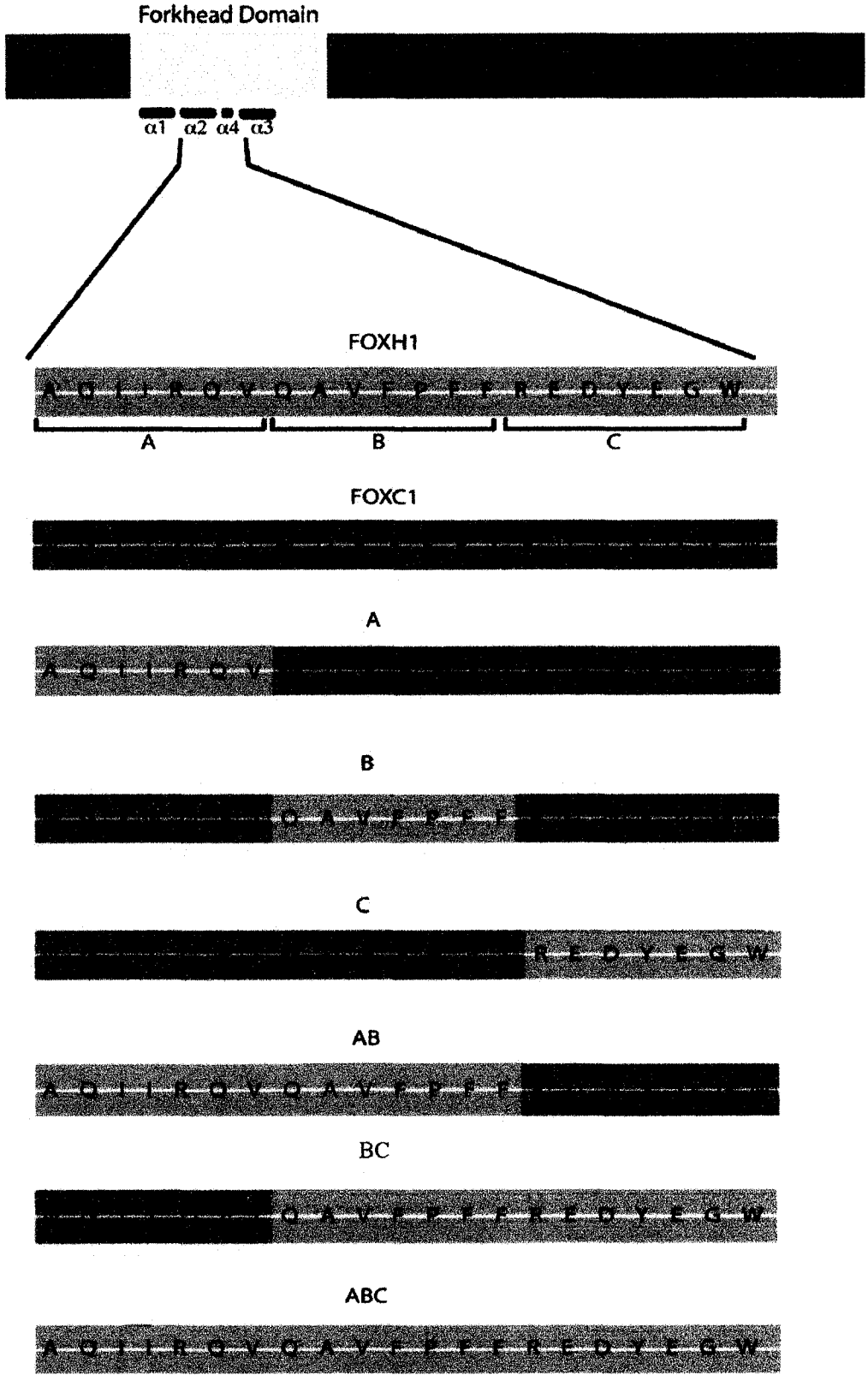
In agreement with previous data, I determined that residues upstream of helix 3 are able to influence DNA-binding specificity. From a region of twenty-one amino acids, I propose that seven residues are key residues that play a more direct role in conferring binding specificity. Homologous transcription factors have highly

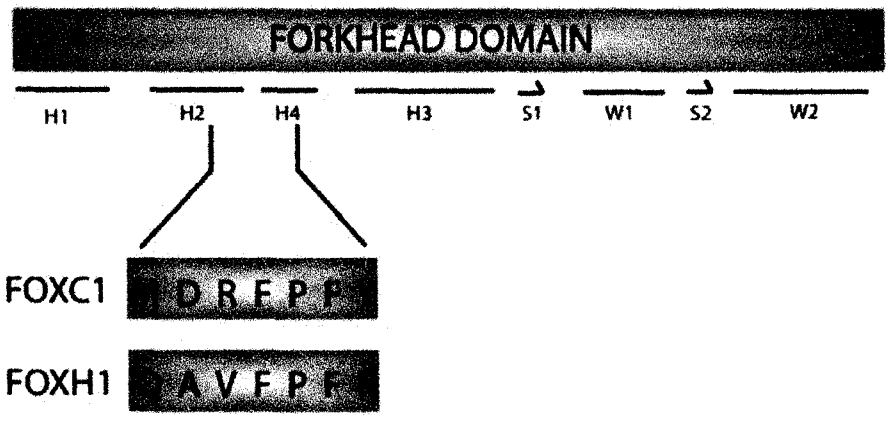
A

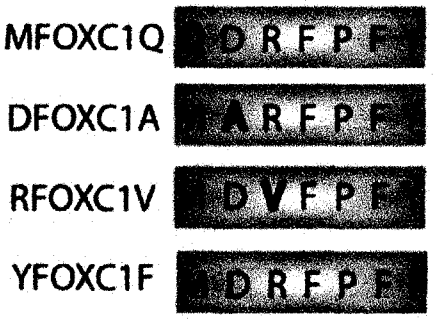


B

FHD CORE CONSENSUS	RYMAAYA
FOXC1 DERIVED BINDING SITE	GTAAATAAA
FOXH1 DERIVED BINDING SITE	CAATACACA







conserved DNA-binding regions, yet the binding specificity and core binding sequence vary significantly. An understanding of the basis for sequence-specific binding is therefore necessary for a comprehensive understanding of how the FOX family of proteins function.

Methods

Domain swap construct construction

FOXC1 was mutated in pGEM T Easy using Quickchange™ mutagenesis kit (Stratagene) and appropriate primers (Appendix A) following the manufacturer's protocol, with the addition of 5% DMSO. Mutagenesis products containing the correct changes were selected and potential mutant constructs were sequenced using an automated sequencer (Li-COR). The fragment containing the correct change(s) was then subcloned into pcDNA4A His/Max by way of *Apa I- RsrII* digest and sequenced manually to confirm. Domain swap constructs A, B, and C were built first. Constructs AB and AC were built using construct A as a template for mutagenesis.

Protein extraction and western analysis

See Methods, Chapter One, pages 55

Eletrophoretic mobility shift assays (EMSAs)

See Methods, Chapter One, page 56

Protein modeling

The structure of FOXC2 was used as a model for determining FOXC1 electrostatic potentials. Residues E97 and E117 of FOXC2 were changed to D97 and D117 so that the FOXC2 sequence used as a template model was identical to FOXC1. The model was loaded into the Swiss-PbdViewer program and the molecular surface was calculated and colored according to the Coulomb calculation of electrostatic potentials.

Results

Creating FOXC1-FOXH1 domain swap constructs

PCR based site-directed mutagenesis was employed to introduce various segments of FOXH1 sequence into the reciprocal location in the FOXC1 cDNA (figure 2-3 and figure 2-4).

Expression of recombinant FOXC1

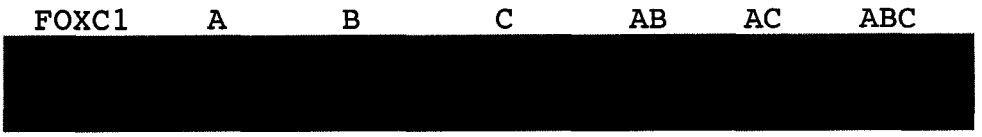
All FOXC1 recombinant constructs were transfected into COS-7 cells and whole cell extracts were resolved by SDS-PAGE and western analysis. Protein was detected by use of an N-terminal (vector encoded) Xpress epitope that demonstrated a product of approximately 80 kDa in size (figure 2-5 A-C). In three independent attempts construct RFOXC1V did not produce a protein of the correct molecular weight and size, but rather a degradation product, and was therefore not tested against the panel of variant oligonucleotides. It is possible that the single substitution from Arginine to Valine renders FOXC1 unstable and unable to produce protein. As this construct did not produce a proper full length protein it was not included in the binding assays.

Analysis of DNA-binding capacity and specificity

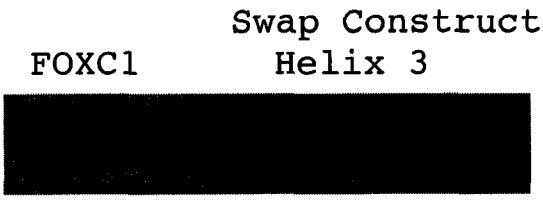
Affinity for the FOXC1 binding site varies among FOXC1/FOXH1 swap constructs

EMSAs were performed as described earlier, see page 65. Swap constructs A, B, AB, and AC have a strong overall affinity for the *in vitro* derived FOXC1 binding site that is comparable to wild type levels (figure 2-6A). Swap construct C displayed a reduction in binding, with an approximate 2-fold reduction in binding affinity compared to wild type levels (figure 2-6A). Swap construct ABC showed a complete loss of binding for the FOXC1 binding site (figure 2-6A) as did the construct containing helix 3 of FOXH1 (figure 6B).

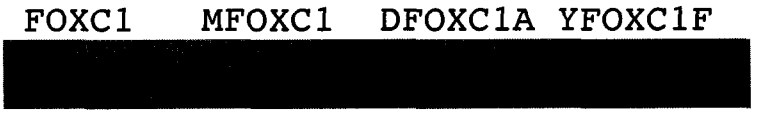
A



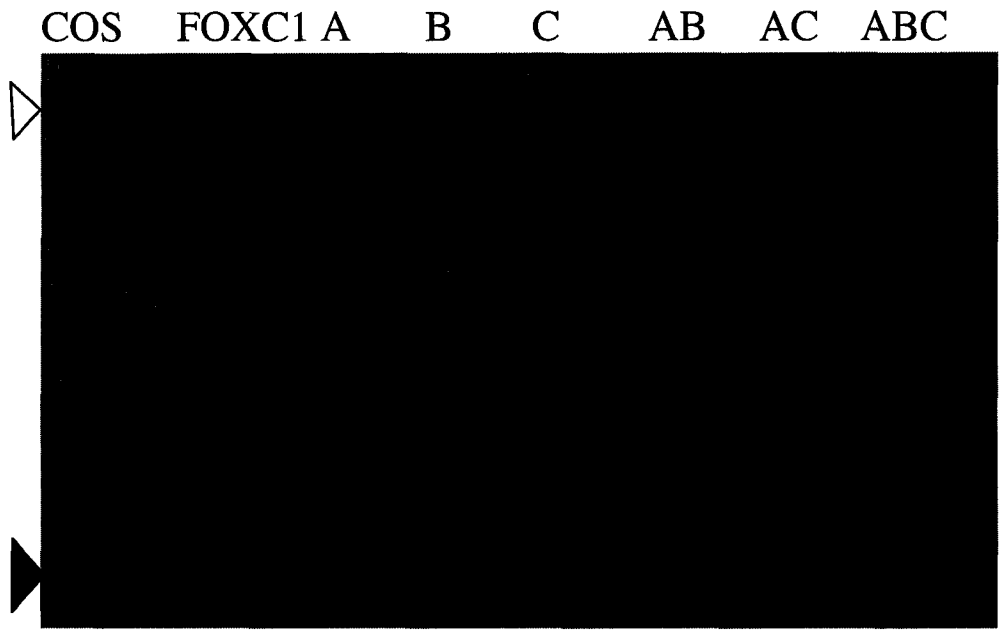
B



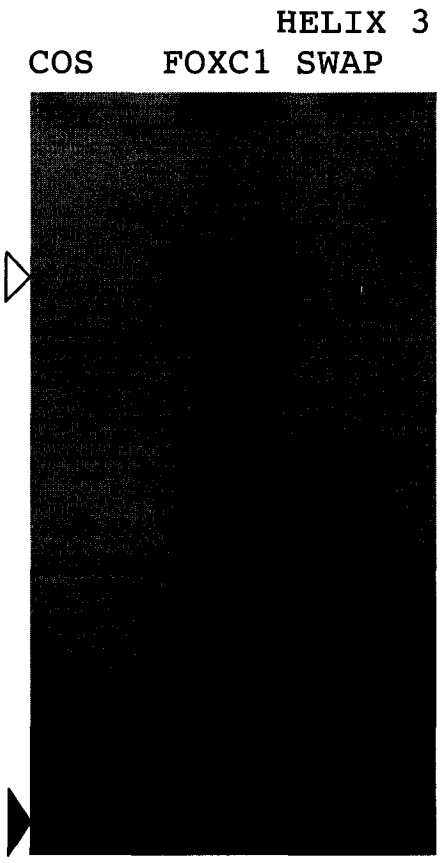
C



A



B

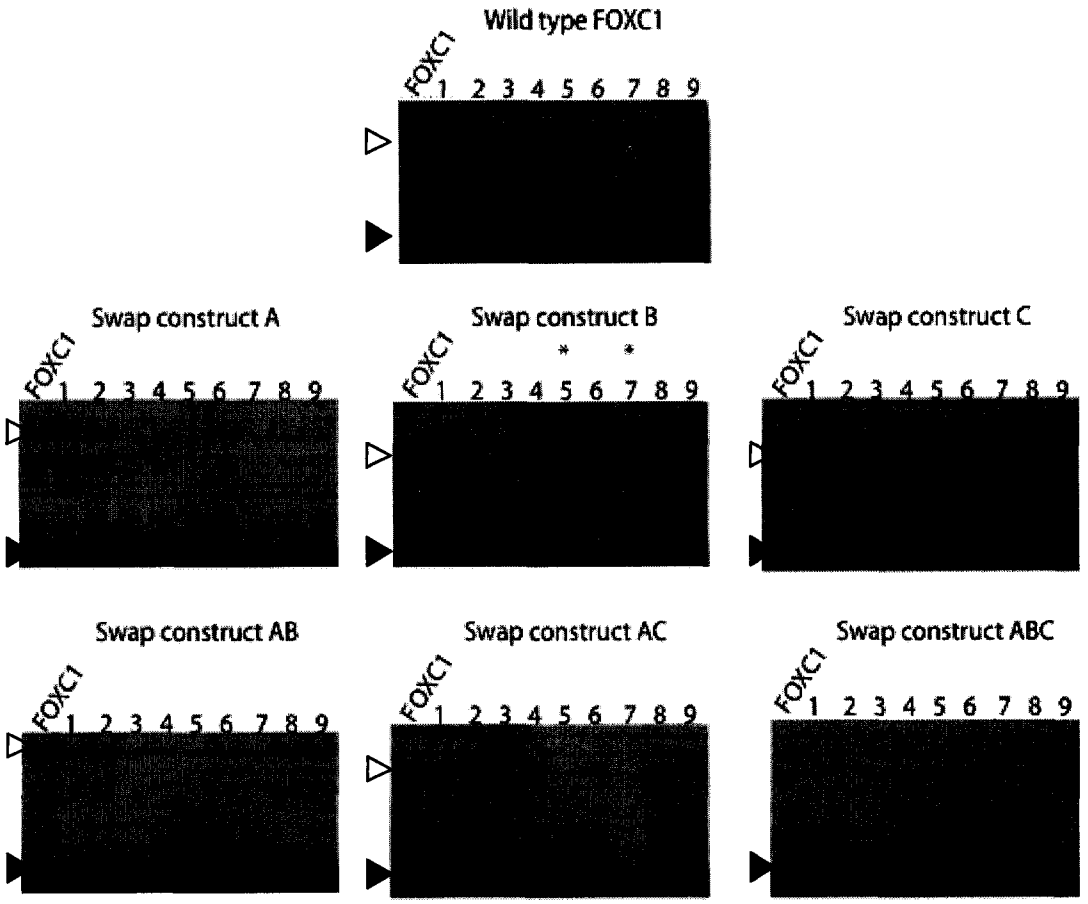


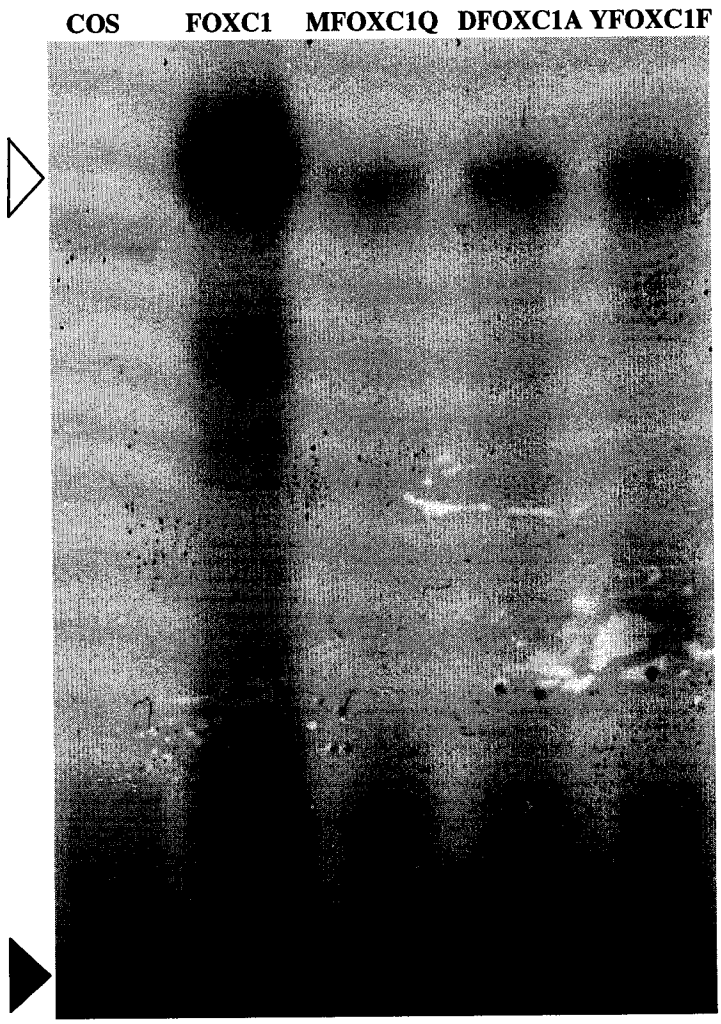
Seven amino acids that precede alpha helix 3 may alter DNA-binding specificity

DNA-binding assays were preformed as previously described on page 65. Swap constructs A, C, AB, and AC show variation in affinities for the variant oligonucleotides in the sense that affinity for all sites is reduced compared to wild type FOXC1 affinities (figure 2-7). While construct A maintains binding ability for sites 8 (alteration to preferred FOXC1 site shown in bold and underlined; GTAAATA**T**A) and 9 (GTAAATAA**T**), affinity for these sites is reduced compared to wild type FOXC1 (figure 2-7). Construct C shows comparable levels of affinity for the wild type FOXC1 site but has reduced affinity for sites 1 (**C**TAAATAAA), 2 (G**A**AAATAAA) 8 (GTAAATA**T**A) and 9 (GTAAATAA**T**). While swap constructs A, C, AB and AC express a general reduction in affinity for the various binding sites, swap construct B is the only construct to exhibit altered DNA-binding specificity by having an increase in affinity for sites otherwise weakly bound by wild type FOXC1 (figure 2-7). The affinity of swap construct B for the variant FOXC1 binding site 5 (GTAA**T**TAAA), and 7 (GTAAAT**T**AA) is much higher than the affinity of the wild type FOXC1 protein for these binding oligonucleotides. Swap constructs ABC, and the construct containing FOXH1 helix 3 did not bind the FOXC1 binding site or any of the variant sites.

Single amino acid changes in construct B can disrupt FOXC1 DNA-binding ability

Swap construct B (which contains 4/7 swapped residues) maintains near wild type levels of DNA-binding (figure 2-6) for the consensus site, however, this construct indicates altered specificity for the panel of variant binding sites (figure 2-7). To further investigate the amino acids in this region, single amino acids were altered separately and analyzed for alterations in binding both the consensus FOXC1 site and the variant sites. Altering a specific amino acid within this region is sufficient to reduce FOXC1 affinity for its *in vitro* derived preferred binding site. MFOXC1Q, DFOXC1A, and YFOXC1F all have reduced levels of binding comparable to wild type levels (figure 2-8). Therefore, substitution at these sites can influence affinity for the FOXC1 preferred binding site.



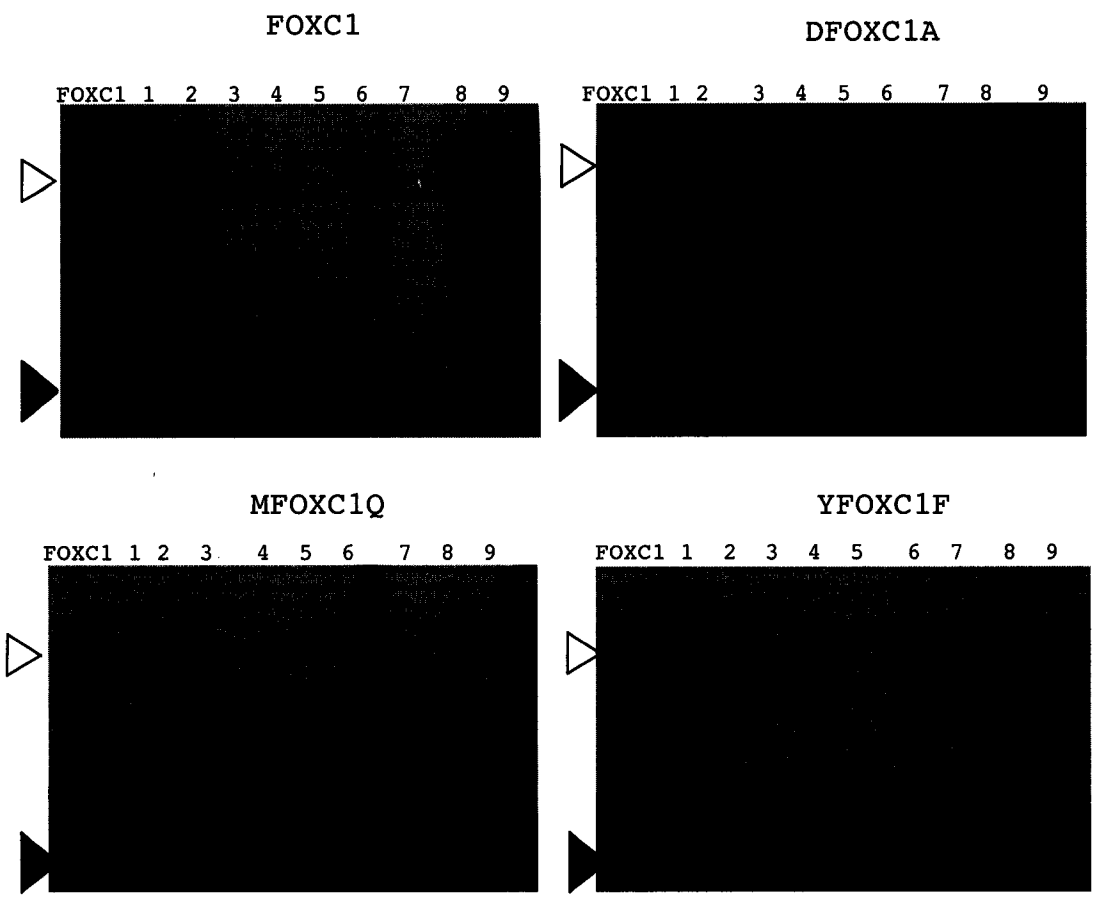


Single amino acid changes do not replicate construct B's altered DNA-binding specificity

Individual swap constructs MFOXC1Q, DFOXC1A, and YFOXC1F do not replicate the change in affinity for sites 5 (GTAA**T**TAAA) and 7 (GTAAAT**T**AA) that were seen when multiple mutations were present in construct B (figure 2-9). Individual changes in the region of construct B result in reduced affinity for some of the variant oligonucleotides. Both DFOXC1A and YFOXC1F show comparable affinities for the wild type FOXC1 site and variant site 1 (**C**TAAATAAA), but have reduced affinity for variant site 2 (G**A**AAATAAA) (figure 2-9). Therefore, while single amino acid changes do not replicate the changes in sequence recognition seen when multiple mutations in construct B are present, the single changes are sufficient to reduce affinities for the FOXC1 preferred site.

Residues in domain swap construct B alter the electrostatic potential of FOXC1

Computer modeling was done by Tim Footz. A computational model of the electrostatic surface potential of FOXC1 and swap construct B indicates that introduction of FOXH1 sequence into the region encompassed by construct B alters the electrostatic potential of FOXC1. As figure 2-10 indicates, the introduction of arginine (DFOXC1A) results in loss of a negative surface charge that is present in wild type FOXC1. However, as indicated above, while this alteration alone is able to reduce the affinity for binding the FOXC1 preferred site, it is not able to replicate the increased affinity for variant sites weakly bound by FOXC1.



Discussion

Analysis of FOXC1 DNA-binding specificity

Trying to make predictions about the mechanism FOXC1 uses to distinguish one sequence from another is difficult in the absence of structural data for FOXC1. With existing models of homologous proteins it is possible to extrapolate structural predictions and make suggestions as to the region employed by FOXC1 to recognize specific sequences. In turn, a working hypothesis can be made regarding the possible mechanism(s) used by FOXC1 to discriminate between sequences.

Analysis of the region preceding helix 3

One year after the x-ray crystal structure of FOXA3 was solved, it was suggested that the twenty amino acids preceding the recognition helix H3, were the primary controlling element for determining DNA-binding specificity (58) (61). As the structures of other forkhead proteins were solved, the implications for this region of amino acids to play a role in conferring DNA-binding specificity strengthened (56) (63) (72).

Amino acid changes were introduced into FOXC1 in an attempt to determine if FOXC1 binding specificity was governed by the same conventions as other forkhead proteins. When FOXC1 swap constructs were tested in binding assays using the *in vitro* derived FOXC1 binding site, constructs A, B, AB, and AC showed levels of binding comparable to wild type, while construct C indicated reduced binding (figure 2-6). Therefore some changes in FOXC1 sequence preceding helix 3 still render recombinant FOXC1 able to recognize and bind its determined binding site at near wild type levels. Construct ABC did not retain any ability to bind the FOXC1 binding site, suggesting residues in this region play a prominent role in DNA-binding and/or sequence recognition. Consistent with my observations, in the domain swap experiments undertaken by Overdier *et al.,* (61) swapping the same region of twenty amino acids of FOXA3 for the reciprocal amino acids of FOXQ1 rendered FOXA3 unable to recognize its specific target sequence, and instead bind the sequence specific for FOXQ1. It is possible that the loss of binding with construct ABC would

be recapitulated as an increase in affinity for the FOXH1 binding site, however due to the suboptimal recognition of FOXH1 for its own binding site, such an obvious switch in recognition of binding sites could not be confirmed.

Domain swap constructs were subjected to binding assays against a panel of variant FOXC1 binding sites (Table 1-1). Of the six swap constructs tested, only construct B, through an increase in affinity for sites otherwise weakly bound by wild type FOXC1, showed a significant divergence in binding pattern for the variant oligonucleotides (figure 2-7). Swap construct B resides in the middle section of the total twenty amino acid region that is predicted to play a role in binding specificity (63) (61). In their investigation of forkhead proteins, Pierrou *et al.* (63) suggest that preference for position +6 of the core consensus site (RYMAA**Y**A, position +6 in boldface) is determined by a region in the central part of the forkhead domain. The area encompassed by swap construct B corresponds to the identical region described by Pierrou *et al.* (63) adding support for this region being a primary controlling element in determining DNA-binding specificity in forkhead proteins.

Construct B comprises a seven amino acid stretch with four of the seven residues differing between FOXC1 and FOXH1. Construct B demonstrated altered sequence recognition abilities therefore the role this region plays in determining binding specificity was further investigated by creating constructs that differed by only a single amino acid (figure 2-4, summary table 2-1). The constructs, MFOXC1Q, DFOXC1A, and YFOXC1F, while they did have an overall reduction in affinity for all sites, did not replicate the alteration in binding affinities shown by construct B (figure 2-9). This does not contradict previous evidence suggesting that this region is involved in binding specificity, but rather suggests that these single amino acid changes are not sufficient to confer altered binding specificity. Furthermore, the reduction in affinity for the variant sites suggests that at some level, these residues are involved in maintaining proper levels of DNA-binding. Studies with the insect protein Fork head, suggest that the Fork head-DNA interaction is context dependent and that the binding specificity of the protein is partly determined by specific combinations of neighboring bases (73). Certain dinucleotide steps (combinations of neighbouring bases) display a profound effect on the Fork head

Table 2-1. Summary of binding defects of FOXC1/ FOXH1 swap constructs, single amino acid swap constructs and FOXH1 helix 3 swap construct.

Swap Construct	DNA binding capacity	Altered binding specificity
A	+++	NO
B	+++	YES
C	++	NO
AB	+++	NO
AC	+++	NO
ABC	-	NO
Helix 3	-	NO
MFOXC1Q	++	NO
DFOXC1A	++	NO
YFOXC1F	++	NO

DNA-binding scoring: (+++) 81-99%; (++) 51-80%; (+) 21-50%; (+/-) 11-20%; (-) 0-10%

ability to bind DNA, and the presence of these base combinations significantly adds to the overall specificity of the Fork head binding (73). A base combination alters various parameters of the DNA molecule and thus it can potentially affect its ability to bind proteins. This suggests, that while single amino acid changes in the FOXC1 FHD may not be sufficient to confer DNA-binding specificity, di-nucleotide changes could be sufficient.

What has been accomplished is a paring down of the twenty amino acids initially suggested to be involved in sequence recognition, to a region of seven residues that can alter DNA recognition properties. While analysis of my results cannot determine the residues *necessary* to control binding specificity, they do determine that a group of seven residues within the initial twenty in question, are indeed *sufficient* to change the binding preference of FOXC1.

Analysis of the validity of homology modeling

FOXA3 was the first forkhead protein to have its structure solved (58). DNA footprinting studies as well as site selection studies both determined a core consensus binding site for forkhead proteins (60) (63). As mentioned, a 13 bp oligonucleotide was used in determining the structure of FOXA3. This oligonucleotide was based on the sequence determined from a FOXA3 DNA footprint and TTR promoter analysis. Figure 2-11 illustrates the sequences of the defined core consensus, the FOXA3 footprint, the TTR promoter and the sequence used in the x-ray crystallography of FOXA3 (58). In the oligonucleotide used for determining the structure of FOXA3, positions +6 and +7 of the core consensus do not correspond to the consensus sequence or the TTR promoter sequence. The T at position +6 and A at position +7 of the TTR promoter have been altered to a C and C in the oligonucleotide used for determining the crystal structure (figure 2-11). This has caused some speculation as to the validity of the determined FOXA3 structure (63) since numerous contacts are made at these positions by α-helices 1 and 3 and wing 1 (figure i-11) (58). Residues outside the core consensus, that is sequences flanking the core, have been shown to influence binding (63) therefore it is possible that alteration of these residues in the core sequence may also disrupt the default binding interactions.

Core consensus		**RYMAAYA** (+6)
FOXA3 Footprint	5'	CTAAGTCAATA 3'
	3'	GATTCAGTTAT 5'
TTR Promoter	5'	CTAAGTCAATAATCAG 3'
	3'	GATTCAGTTATTAGTC 5'
Oligonucleotide for FOXA3 crystal	5'	GACTAAGTCAACC 3'
	3'	CTGATTCAGTTGG 5'

Additional concern for the validity of the FOXA3 structure stems from the length of the oligonucleotide used to determine the structure. When the NMR structures of FOXD3 and FOXO4 were determined, 17mer oligonucleotides were used (62) (64). The 13mer oligonucleotide used in determining the structure of FOXA3 does not extend beyond the 3' border of the core. It is possible that contacts between DNA and *FOXA3* in the 3' region are suboptimal and that conceivable interactions or distortions in the DNA are not able to be detected with this short oligonucleotide (63). Bending assays with the FHD of FOXC1 and FOXD1 concluded that the proteins bend the DNA at an angle of 80° and 90°, respectively. However, bending assays with the entire FOXC1 protein indicate that in fact, *FOXC1* bends the DNA at an angle of 112° (55). In the x-ray crystal structure with FOXA3, the DNA has a curvature of only 13° (58). This discrepancy either reflects a difference in binding characteristics or inconsistencies in experimental methods.

With these concerns in mind, I suggest the x-ray crystal structure of FOXA3 nevertheless does provide a good structural model that can be used to extrapolate predictions about other forkhead proteins. With the exception of the different orientation of FOXD3 helix 3, superimposing the NMR structures of FOXD3, FOXO4 and FOXC2 does not show gross structural differences (figure i-15 b). The DNA-binding domain of forkhead proteins is highly homologous among family members, and it is likely that proteins from this family employ similar tactics to determine sequence specificity. Therefore, gross differences in the determined structures of FOX proteins could indicate either variations in binding mechanisms or inconsistencies with structure determination. It is the absence of significant structural differences between the known FOX protein-DNA complexes that I feel asserts FOXA3 as a good structural model for other forkhead proteins.

Mechanisms of DNA-binding specificity

While structural data from various forkhead proteins support the hypothesis that the amino acids preceding helix 3 are involved in DNA-binding specificity, the mechanism by which this region confers specificity is yet to be determined and remains controversial. In examining a stereo diagram of the backbone C^{α} atom

coordinates of the core helices of FOXD3 and FOXA3, it is clearly visible that the major differences lie in the C-terminal part of helix 3 (figure i-13 b). The formation of a small fourth helix H4 is suggested to be the mechanism by which DNA-binding specificity is achieved for FOXD3 (56). This hypothesis has come into question with the addition of the NMR structure of FOXO4, which although more distantly related to FOXA3, nevertheless shares remarkable structural homology to the FOXA3 forkhead domain (64) (58). Weigelt *et al.* suggest that the presence of helix 4 is not uncommon to forkhead proteins and propose that changes in electrostatic potential may alter interactions of the recognition helix with the major groove of DNA and thereby confer binding specificity (64).

Computational analysis of the electrostatic potential of FOXC1 and domain swap construct B indicate that changes in surface charge may influence the ability of FOXC1 to recognize specific sequences. As figure 2-10 indicates, introduction of different residues into the region encompassed by swap construct B, results in loss of a negative surface charge that is present in wild type FOXC1. Reduction in negative charges could potentially result in greater forces of attraction to the DNA, thereby shifting the position of residues in the immediate vicinity, as well as entire sub-domains, especially the neighboring recognition helix H3. Computer modeling indicates there is a difference in the surface potential of FOXC1 and swap construct B, and my binding assays indicate that swap construct B has altered sequence preferences compared to wild type FOXC1 (figure 2-7). Therefore on the basis of my data, I suggest that changes in electrostatic potential can influence the sequence recognition properties of FOXC1.

A recent study provides further evidence of electrostatic changes influencing a protein's ability to recognize specific sequences. The effect of charge changes on key residues of FOXC1, were investigated by converting each residue to an alanine (neutral charge), glutamic acid (negative charge), or lysine (positive charge) residue (74). Prominent effects in binding affinity were seen with residues I126 and R127. When converted to alanine or glutamic acid, binding capacity for residue I126 is completely disrupted. Residues R127A and R127K show minimal amounts of binding, however levels of capacity are reduced well below 10X wild type levels. The

R127E conversion shows no capacity for binding (74). Of greater importance is the observation that R127A and R127K have significantly different preferences for the variant FOXC1 binding sites. These results indicate that changing the charge of residues in the N-terminal portion of helix 3 can alter the sequence discrimination ability of FOXC1. These data, in conjunction with my computational analysis, provide further evidence for electrostatic potential playing a key role in sequence recognition.

Clearly there are a number of forces controlling how forkhead proteins recognize different sequences. With previous investigations and the data reported herein, it is clear that the region preceding helix 3 plays a role in orchestrating DNA binding and sequence recognition. My data support the concept that differences in electrostatic potential have possible influences on DNA binding. A working hypothesis regarding the mechanism used by FOX proteins to determine sequence recognition can be made from my analysis of FOXC1. My results clearly indicate that the sequence preceding helix 3, more specifically seven residues in the central region of the FHD, influence the sequence recognition properties of FOXC1. Previous work concerning charge properties of key residues, and our computational model of surface charges, suggest that electrostatic potential of the binding interface has a role in controlling sequence recognition ability (74). Therefore, based on my results and in conjunction with recent studies, I hypothesis that a combination of forces control the ability of FOXC1 to discriminate one sequence from another. I believe a combination of indirect interactions with the DNA made by specific sequences preceding helix 3 and the charge potential of the binding interface are both key controlling elements in conferring sequence recognition abilities. It is likely that this hypothesis can be extended to other homologous FOX family members. DNA bending and other, as of yet undiscovered factors, may also play a role in determining DNA-binding specificity and further investigations in this area need to be considered.

General Discussion and Conclusions

The two lines of investigation discussed herein have dealt with both the functional as well as structural aspects of *FOXC1* with the intent to gain a further appreciation for the molecular basis of Axenfeld-Rieger malformations. Analyses of missense mutations in wing 2 have proven the importance of this region for FOXC1 function in reference to DNA binding and transactivation ability. Investigations into the residues involved in specific sequence recognition have highlighted a region in the central part of the FOXC1 FHD as a key controlling element.

This analysis of missense mutations was undertaken to gain an understanding of FOXC1 function using biologically relevant disease-causing mutations. In conjunction with analyzing FOXC1 function, homology modeling was used to extrapolate structural implications regarding protein mutation. To these ends, it was found that missense mutations in wing 2 of the FOXC1 FHD are able to perturb FOXC1 function. The G165R mutation does not face the binding interface, and therefore does not contribute to the make up of the hydrophobic pocket (figure 1-9). As such, this mutation does not disrupt the ability of FOXC1 to bind DNA, however, G165R does perturb the ability of FOXC1 to transactivate a reporter construct (figure1-8), suggesting a role for wing 2 in intermolecular interactions. Mutations that are involved in maintaining the hydrophobic core, such as M161K, were shown to disrupt the ability of FOXC1 to bind DNA. On the basis of homology modeling, I suggest this is a result of disruptions to local interactions with neighboring residues (figure 1-10). A loss of binding capacity with this mutation is reflected in a similar loss in transactivation ability (figure 1-8). R169P, the sole base contact in the minor groove, clearly disrupts the ability of FOXC1 to bind the consensus binding site (figure 1-6). These mutation analyses have demonstrated both the functional significance of individual amino acids in the FHD, as well as the importance of wing 2 as a secondary structure in the FHD.

A second aim of my investigation into FOXC1 function was to determine the residues important for conferring DNA-binding specificity in FOXC1. A domain swap approach was used to investigate whether the amino acids that precede helix 3 are involved in conferring sequence discrimination ability. My results suggest that a region of seven amino acids in the region intervening between α-helix 2 and α-helix

3 play an important role in sequence recognition. Single amino acid changes are not sufficient to alter sequence preferences, however changing four out of seven residues is sufficient to change the affinity of FOXC1 for a series of variant FOXC1 binding sites. While the precise residues that are responsible for controlling binding specificity could not be determined, this work provides sufficient information to continue investigating the controlling element utilized by FOXC1, and other FHD proteins, to distinguish between different sequences. While at present, a structure of FOXC1 complexed with DNA is unavailable, the structures of homologous FOX proteins can be used as a working model. Comparing FOXC1 with these known structures suggests that the family of forkhead transcription factors relies upon the same general region of amino acids to help control sequence recognition.

These investigations, both into the molecular disturbances caused by missense mutations as well as trying to determine the main controlling mechanism behind sequence recognition, have been valuable not only for their individual insights into separate functions of the FHD, but into the domain as a functioning unit. Except for the studies described herein, no other analyses have been done on FOXC1 to test the effects of missense mutations in the wing 2 region or to analyze in detail the FHD in regards to specific amino acids involved in binding specificity. Computational modeling has proven helpful in making predictions regarding alterations to protein structure and determining key residues (54) (55), and until the structure of FOXC1 is determined, this will continue to be a valuable method for enhancing biochemical investigations.

Analysis of the role of wing 2 in DNA-binding specificity

While numerous studies have implicated the central region of the forkhead domain as being the primary controlling element for DNA-binding specificity, several studies have alluded to the involvement of wing 2 in sequence discrimination. Wing 2 of FOXA3 contacts the DNA backbone in the minor groove on the 5' side of the core consensus site, while wing 1 navigates closer to the 3' end (58). However, it is sequence in wing 2 that influences binding by affecting the 3' flanking sequence of the core consensus site (63). Pierrou *et al.* (63) suggest that wing 2 influences DNA-

protein interactions 3' of the core consensus by indirect influences or that residues outside of the C-terminal portion of the FHD contribute to determining specificity.

My own analysis of missense mutations in wing 2 indicates that while indirect involvement of wing 2 in binding specificity remains inconclusive, wing 2 is necessary for direct interactions with the DNA to maintain proper binding. Wing 2 meanders across the surface of the three helix bundle and aromatic residues in wing 2 interact with other residues in the bundle, contributing to the stability of wing 2 (57). The creation of temperature-sensitive wing 2 mutants demonstrated that stabilization of wing 2 is critical for DNA binding activity and therefore the transcriptional function of the protein (57). Shiyanova and Liao suggest the amino acid sequence in the wings influence the dissociation rate of the protein-DNA complex and suggest that association rates may have an influence on sequence discrimination (75).

Whether it results from the formation of a fourth helix or through differences in electrostatic potential, orientation of the recognition helix with respect to the DNA, is the prominent working hypothesis in how forkhead proteins confer sequence discrimination. Analysis of the FOXC2/FOXD3 distance difference matrix suggests that the position of the wings are correlated to the position of the recognition helix (72). This implies that the orientation of helix 3 may influence which residues in wing 2 may be involved in indirect interactions with the DNA and contribute to binding specificity. One structural hypothesis is that helices 1 and 2 provide a solid structural framework that supports a flexible helix 3. The region between helix 4 and 3 acts similar to a hinge to present the recognition helix in an optimal position. This leaves the wings to provide definite interactions to the DNA that either further stabilize the complex or enhance the orientation of helix 3, thereby affecting sequence recognition (72). The position of helix 3 relative to other secondary elements that make DNA contacts, such as wing 1 and wing 2, could fine tune the final DNA contact scheme used by each forkhead protein (70).

When the winged-helix-turn-helix domain was first defined it was suggested the wings were situated similar to those of the wings of a butterfly, hence the term 'winged helix'. Bending assays, both with the FHD of FOX proteins and full length FOXC1 suggests that bending of the target site is an intrinsic characteristic of the

forkhead family of proteins (55) (63). If indeed, bending of the DNA is an intrinsic characteristic of this family of proteins, it suggests a new role for the wings of the butterfly. Instead of acting as peripheral wings, perhaps the wings wrap around and warp the DNA, inducing it to bend and thereby causing specific residues to interact with the DNA.

With several working hypotheses for the role wing 2 plays in sequence recognition, it is impossible to determine which is the more accurate. Based on both my analysis of missense mutations in wing 2 and on the controlling element for FOXC1 binding specificity, I have come to conclude that wing 2 does play a role in helping FOX proteins bind specific sequences, but that this role is more peripheral than direct. Structural data show that wing 2 contacts the DNA in the minor groove (58). This direct relationship between protein and DNA places wing 2 in a position where numerous indirect contacts could be made. Altering M161 and G165, which do not have direct contact with the DNA, is sufficient to disrupt binding ability. While these same mutations alone, are not sufficient to change the sequence preference of the protein, the loss of binding affinity suggests that at some level, residues in wing 2 are influencing the binding properties of FOX proteins. I hypothesis that wing 2 contains residues that, while not necessary for specific sequence recognition, do contribute to the overall process of sequence recognition.

Future Directions

A tremendous amount of work has been done to understand FOXC1, both developmentally and biologically. Mutation analyses have allowed functional roles to be assigned to sub-domains of the FHD, correlating function with structure (figure 1-11). Computational analyses have aided in visualizing alterations to protein structure that help explain functional losses (figure 1-10). However, while these functional assignments to the sub-domains are helpful, they remain incomplete. With respect to the FHD, the question of controlling DNA-binding specificity remains a major source of interest. The least number of amino acids necessary to effect a change in specificity has not been determined. Two major questions regarding binding specificity remain unanswered: 1) What is the minimum number of residues that are necessary to confer DNA-binding specificity and 2) What is the mechanism behind sequence discrimination? My research has refined the initial region of twenty amino acids believed to be involved in binding specificity, to a segment of seven amino acids that are sufficient to effect a change in the sequence recognition ability of FOXC1. Site selection and amplification has been performed on FOXC1 (63) and a consensus binding site determined for this protein. Swap construct B has shown significant changes in affinity for the variant FOXC1 binding sites, therefore site selection and amplification with this construct may reveal a better understanding of the residues involved in discriminating against specific sequences.

Without a determined structure of FOXC1 it is difficult to know how residues behave in the presence of different DNA sequences. Determining the structure of FOXC1, complexed to the preferred FOXC1 site, would be the quintessential analysis in determining how FOXC1 functions to recognize specific sequences. Several methods could be used to analyze how FOXC1 behaves towards different sequences. Solving the structure of wild type FOXC1 complexed to DNA and comparing this to a solved structure of mutant FOXC1, perhaps containing residues from construct B, would help delineate how residues that make indirect contacts influence protein-nucleic acid interactions. It would also be beneficial to determine the structure of FOXC1 bound to the preferred FOXC1 binding site as well as to variant sites 5

(GTAATTAAA) or 7 (GTAAATTAA), both of which are weakly bound by FOXC1 but strongly bound by construct B.

From the data presented in my analysis of FOXC1 binding specificity, it remains uncertain as to what is the mechanism controlling sequence recognition. The data reported herein support the hypothesis that the region preceding helix 3 is an important element in controlling DNA binding specificity. The recent data concerned with changes in surface charges present a new angle to investigate the mechanism behind sequence recognition. Therefore, changing the charge of residues in the region preceding helix 3 (for example, swap construct B) may shed some light onto the controversial issue of sequence recognition.

Although my analysis of single missense mutations in wing 2 has proven to cause no alterations in sequence recognition, there is support for involvement of wing 2 in determining binding specificity (63) (57). Whether it is through indirect interactions with the DNA or structural changes such as DNA bending, wing 2 remains an area of interest. Wing 2 is one of the less conserved regions of the FHD, which makes it a good candidate for being involved in protein specific functions. Further analysis into the wing 2 region, and perhaps of the C-terminal region just outside the FHD, may shed some light onto the functional importance of this region in regard to DNA-binding.

Appendix A. Primers used for site directed mutagenesis

Amino Acid Change	*Forward and Reverse primers*
M161K	
Forward, 5' – 3'	ccggactcctacaacaagttcgagaacggcagc
Reverse, 5' – 3'	gctgccgttctcgaacttgttgtaggagtccgg
G165R	
Forward, 5' – 3'	aacatgttcgagaaccgcagcttcctgcggcgg
Reverse, 5' – 3'	ccgccgcaggaagctgcggttctcgaacatgtt
R169P	
Forward, 5' – 3'	gagaacggcagcttcctgccgcggcggcggcgcttcaag
Reverse, 5' – 3'	cttgaagcgccgccgccgcggcaggaagctgccgttctc
Swap A	
Forward, 5' – 3'	ccggacaagaagatcaccctggcccagatcatccgtcaggtcatggaccgcttccccttctac
Reverse, 5' – 3'	gtagaaggggaagcggtccatgacctgacggatgatctgggccagggtgatcttcttgtccgg
Swap B	
Forward, 5' – 3'	aacggcatctaccagttcatccaggccgtgttccccttcttccgggacaacaagcagggctgg
Reverse, 5' – 3'	ccagccctgcttgttgtcccggaagaaggggaacacggcctggatgaactggtagatgccgtt
Swap C	
Forward, 5' – 3'	ttccccttctaccgggaggactatgagggctggcagaacagc
Reverse, 5' – 3'	gctgttctgccagccctcagagtcctcccggtagaaggggaa
Swap AB	
Forward, 5' – 3'	gcccagatcatccgtcaggtccaggccgtgttccccttcttccgggacaacaagaag
Reverse, 5' – 3'	cttcttgttgtcccggaagaaggggaacacggcctggacctgacggatgatctgggc
Swap AC	
Forward, 5' – 3'	atggaccgcttccccttctaccgggaggactatgagggctggcgggacaacaagcag
Reverse, 5' – 3'	ctgcttgttgtcccgccagccctcatagtcctcccggtagaaggggaagcggtccat
Swap ABC	
Forward, 5' – 3'	aagaagatcaccctggcccagatcatccgtcaggtccaggccgtgttcccc
Reverse, 5' – 3'	ggggaacacggcctggacctgacggatgatctgggccagggtgatcttctt
Swap Helix 3	
Forward, 5' – 3'	gacaacaagcagggctggaaggactccatccgccacaacctttcctctaatcgg
Reverse, 5' – 3'	caccttgacgaagcaccgattagaggaaaggttgtggcggatggagtccttcca
MFOXC1Q	
Forward, 5' – 3'	atctaccagttcatccaggaccgcttccccttc
Reverse, 5' – 3'	gaaggggaagcggtcctggatgaactggtagat
DFOXC1A	
Forward, 5' – 3'	taccagttcatcatggcacgcttccccttctac
Reverse, 5' – 3'	gatgaaggggaagcgtgccatgatgaactggta
RFOXC1V	
Forward, 5' – 3'	cagttcatcatggacgtgttccccttctaccgg
Reverse, 5' – 3'	ccggtagaaggggaacacgtccatgatgaactg
YFOXC1F	
Forward, 5' – 3'	gaccgcttccccttcttccgggacaacaagcag
Reverse, 5' – 3'	ctgcttgttgtcccggaagaaggggaagcggtc

Biblogoraphy

1. Quigley, H.A. (1996) Number of people with glaucoma worldwide. *Br J Ophthalmol,* 80, 389-93.
2. Friedman, J.S. and Walter, M.A. (1999) Glaucoma genetics, present and future. *Clin Genet,* 55, 71-9.
3. Jakobiec, F. (1982) *Ocular embryology, anatomy, and teratology.* Harper and Row, Philadelphia.
4. Ritch, R., Shields, B., and Krupin, T. (1989) *The Glaucomas.* the C.V. Mosby Company, St. Louis.
5. Axenfeld, T.H. (1920) Embryotoxon cornea posterius. *Klin. Monastbl. Augenheilkd.,* 65, 381-382.
6. Alward, W.L. (2000) Axenfeld-Rieger syndrome in the age of molecular genetics. *Am J Ophthalmol,* 130, 107-15.
7. Rieger, H. (1935) Beitrage zur kenntnis seltener missbildungen der iris. II. Uber hypoplasie der irisvorderblattes mit verlagerung und entrundung der pupille. *Graefe. Arch Klin Exp Ophthalmol,* 133, 602-635.
8. Kulak, S.C., Kozlowski, K., Semina, E.V., Pearce, W.G. and Walter, M.A. (1998) Mutation in the RIEG1 gene in patients with iridogoniodysgenesis syndrome. *Hum Mol Genet,* 7, 1113-7.
9. Tripathi, B.J. and Tripathi, R.C. (1989) Neural crest origin of human trabecular meshwork and its implications for the pathogenesis of glaucoma. *Am J Ophthalmol,* 107, 583-90.
10. Shields, M.B. (1983) Axenfeld-Rieger syndrome: a theory of mechanism and distinctions from the iridocorneal endothelial syndrome. *Trans Am Ophthalmol Soc,* 81, 736-84.
11. Lines, M.A., Kozlowski, K. and Walter, M.A. (2002) Molecular genetics of Axenfeld-Rieger malformations. *Hum Mol Genet,* 11, 1177-84.
12. Kaiser-Kupfer, M.I. (1989) Neural crest origin of trabecular meshwork cells and other structures of the anterior chamber. *Am J Ophthalmol,* 107, 671-2.
13. Semina, E.V., Reiter, R., Leysens, N.J., Alward, W.L., Small, K.W., Datson, N.A., Siegel-Bartelt, J., Bierke-Nelson, D., Bitoun, P., Zabel, B.U. *et al.* (1996) Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. *Nat Genet,* 14, 392-9.
14. Brooks, B.P., Moroi, S.E., Downs, C.A., Wiltse, S., Othman, M.I., Semina, E.V. and Richards, J.E. (2004) A novel mutation in the PITX2 gene in a family with Axenfeld-Rieger syndrome. *Ophthalmic Genet,* 25, 57-62.
15. Lines, M.A., Kozlowski, K., Kulak, S.C., Allingham, R.R., Heon, E., Ritch, R., Levin, A.V., Shields, M.B., Damji, K.F., Newlin, A. *et al.* (2004) Characterization and prevalence of PITX2 microdeletions and mutations in Axenfeld-Rieger malformations. *Invest Ophthalmol Vis Sci,* 45, 828-33.
16. Phillips, J.C. (2002) Four novel mutations in the PITX2 gene in patients with Axenfeld-Rieger syndrome. *Ophthalmic Res,* 34, 324-6.

17. Wang, Y., Zhao, H., Zhang, X. and Feng, H. (2003) Novel Identification of a Four-base-pair Deletion Mutation in PITX2 in a Rieger Syndrome Family. *J Dent Res,* 82, 1008-12.
18. Kozlowski, K. and Walter, M.A. (2000) Variation in residual PITX2 activity underlies the phenotypic spectrum of anterior segment developmental disorders. *Hum Mol Genet,* 9, 2131-9.
19. Semina, E.V., Datson, N.A., Leysens, N.J., Zabel, B.U., Carey, J.C., Bell, G.I., Bitoun, P., Lindgren, C., Stevenson, T., Frants, R.R. *et al.* (1996) Exclusion of epidermal growth factor and high-resolution physical mapping across the Rieger syndrome locus. *Am J Hum Genet,* 59, 1288-96.
20. Gage, P.J. and Camper, S.A. (1997) Pituitary homeobox 2, a novel member of the bicoid-related family of homeobox genes, is a potential regulator of anterior structure formation. *Hum Mol Genet,* 6, 457-64.
21. Gage, P.J., Suh, H. and Camper, S.A. (1999) Dosage requirement of Pitx2 for development of multiple organs. *Development,* 126, 4643-51.
22. 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, 140-7.
23. Lai, E., Prezioso, V.R., Smith, E., Litvin, O., Costa, R.H. and Darnell, J.E., Jr. (1990) HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev,* 4, 1427-36.
24. Weigel, D. and Jackle, H. (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell,* 63, 455-6.
25. Kaufmann, E. and Knochel, W. (1996) Five years on the wings of fork head. *Mech Dev,* 57, 3-20.
26. 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, 1316-28.
27. Hiemisch, H., Monaghan, A.P., Schutz, G. and Kaestner, K.H. (1998) Expression of the mouse Fkh1/Mf1 and Mfh1 genes in late gestation embryos is restricted to mesoderm derivatives. *Mech Dev,* 73, 129-32.
28. Winnier, G.E., Kume, T., Deng, K., Rogers, R., Bundy, J., Raines, C., Walter, M.A., Hogan, B.L. and Conway, S.J. (1999) Roles for the winged helix transcription factors MF1 and MFH1 in cardiovascular development revealed by nonallelic noncomplementation of null alleles. *Dev Biol,* 213, 418-31.
29. Swiderski, R.E., Reiter, R.S., Nishimura, D.Y., Alward, W.L., Kalenak, J.W., Searby, C.S., Stone, E.M., Sheffield, V.C. and Lin, J.J. (1999) Expression of the Mf1 gene in developing mouse hearts: implication in the development of human congenital heart defects. *Dev Dyn,* 216, 16-27.
30. Kume, T., Deng, K. and Hogan, B.L. (2000) Murine forkhead/winged helix genes Foxc1 (Mf1) and Foxc2 (Mfh1) are required for the early organogenesis of the kidney and urinary tract. *Development,* 127, 1387-95.

31. Kume, T., Deng, K.Y., Winfrey, V., Gould, D.B., Walter, M.A. and Hogan, B.L. (1998) The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. *Cell,* 93, 985-96.
32. Topczewska, J.M., Topczewski, J., Solnica-Krezel, L. and Hogan, B.L. (2001) Sequence and expression of zebrafish foxc1a and foxc1b, encoding conserved forkhead/winged helix transcription factors. *Mech Dev,* 100, 343-7.
33. Sasaki, H. and Hogan, B.L. (1993) Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development,* 118, 47-59.
34. Hiemisch, H., Schutz, G. and Kaestner, K.H. (1998) The mouse Fkh1/Mf1 gene: cDNA sequence, chromosomal localization and expression in adult tissues. *Gene,* 220, 77-82.
35. Kume, T., Jiang, H., Topczewska, J.M. and Hogan, B.L. (2001) The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev,* 15, 2470-82.
36. Kidson, S.H., Kume, T., Deng, K., Winfrey, V. and Hogan, B.L. (1999) The forkhead/winged-helix gene, Mf1, is necessary for the normal development of the cornea and formation of the anterior chamber in the mouse eye. *Dev Biol,* 211, 306-22.
37. Ferguson, J.G., Jr. and Hicks, E.L. (1987) Rieger's anomaly and glaucoma associated with partial trisomy 16q. Case report. *Arch Ophthalmol,* 105, 323.
38. Cunningham, E.T., Jr., Eliott, D., Miller, N.R., Maumenee, I.H. and Green, W.R. (1998) Familial Axenfeld-Rieger anomaly, atrial septal defect, and sensorineural hearing loss: a possible new genetic syndrome. *Arch Ophthalmol,* 116, 78-82.
39. 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, 1382-8.
40. Lehmann, O.J., Tuft, S., Brice, G., Smith, R., Blixt, A., Bell, R., Johansson, B., Jordan, T., Hitchings, R.A., Khaw, P.T. *et al.* (2003) Novel anterior segment phenotypes resulting from forkhead gene alterations: evidence for cross-species conservation of function. *Invest Ophthalmol Vis Sci,* 44, 2627-33.
41. Jamieson, R.V., Perveen, R., Kerr, B., Carette, M., Yardley, J., Heon, E., Wirth, M.G., van Heyningen, V., Donnai, D., Munier, F. *et al.* (2002) Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. *Hum Mol Genet,* 11, 33-42.
42. Beebe, D.C. and Coats, J.M. (2000) The lens organizes the anterior segment: specification of neural crest cell differentiation in the avian eye. *Dev Biol,* 220, 424-31.
43. Hong, H.K., Lass, J.H. and Chakravarti, A. (1999) Pleiotropic skeletal and ocular phenotypes of the mouse mutation congenital hydrocephalus (ch/Mf1) arise from a winged helix/forkhead transcriptionfactor gene. *Hum Mol Genet,* 8, 625-37.

44. Smith, R.S., Zabaleta, A., Kume, T., Savinova, O.V., Kidson, S.H., Martin, J.E., Nishimura, D.Y., Alward, W.L., Hogan, B.L. and John, S.W. (2000) Haploinsufficiency of the transcription factors FOXC1 and FOXC2 results in aberrant ocular development. *Hum Mol Genet,* 9, 1021-32.
45. Smith, R.S., Zabaleta, A., Savinova, O.V. and John, S.W. (2001) The mouse anterior chamber angle and trabecular meshwork develop without cell death. *BMC Dev Biol,* 1, 3.
46. Rice, R., Rice, D.P., Olsen, B.R. and Thesleff, I. (2003) Progression of calvarial bone development requires Foxc1 regulation of Msx2 and Alx4. *Dev Biol,* 262, 75-87.
47. Yamagishi, H., Maeda, J., Hu, T., McAnally, J., Conway, S.J., Kume, T., Meyers, E.N., Yamagishi, C. and Srivastava, D. (2003) Tbx1 is regulated by tissue-specific forkhead proteins through a common Sonic hedgehog-responsive enhancer. *Genes Dev,* 17, 269-81.
48. Vitelli, F., Morishima, M., Taddei, I., Lindsay, E.A. and Baldini, A. (2002) Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. *Hum Mol Genet,* 11, 915-22.
49. Libby, R.T., Smith, R.S., Savinova, O.V., Zabaleta, A., Martin, J.E., Gonzalez, F.J. and John, S.W. (2003) Modification of ocular defects in mouse developmental glaucoma models by tyrosinase. *Science,* 299, 1578-81.
50. Takahashi, Y., Koizumi, K., Takagi, A., Kitajima, S., Inoue, T., Koseki, H. and Saga, Y. (2000) Mesp2 initiates somite segmentation through the Notch signalling pathway. *Nat Genet,* 25, 390-6.
51. Hope, I.A. and Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell,* 46, 885-94.
52. Frankel, A.D. and Kim, P.S. (1991) Modular structure of transcription factors: implications for gene regulation. *Cell,* 65, 717-9.
53. Berry, F.B., Saleem, R.A. and Walter, M.A. (2002) FOXC1 transcriptional regulation is mediated by N- and C-terminal activation domains and contains a phosphorylated transcriptional inhibitory domain. *J Biol Chem,* 277, 10292-7.
54. Saleem, R.A., Banerjee-Basu, S., Berry, F.B., Baxevanis, A.D. and Walter, M.A. (2001) Analyses of the effects that disease-causing missense mutations have on the structure and function of the winged-helix protein FOXC1. *Am J Hum Genet,* 68, 627-41.
55. Saleem, R.A., Banerjee-Basu, S., Berry, F.B., Baxevanis, A.D. and Walter, M.A. (2003) Structural and functional analyses of disease-causing missense mutations in the forkhead domain of FOXC1. *Hum Mol Genet,* 12, 2993-3005.
56. Marsden, I., Jin, C. and Liao, X. (1998) Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. *J Mol Biol,* 278, 293-9.
57. Stevens, K., Cirillo, L. and Zaret, K.S. (2000) Creating temperature-sensitive winged helix transcription factors. Amino acids that stabilize the DNA binding domain of HNF3. *J Biol Chem,* 275, 30471-7.

58. 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, 412-20.
59. Murphy, T.C., Saleem, R.A., Footz, T., Ritch, R., McGillivray, B., and Walter, M.A. (2004) The wing 2 region of the FOXC1 Forkhead Domain is necessary for normal DNA-binding and transactivation functions. *Invest Ophthalmol,* 45, 2531-2538.
60. Costa, R.H., Grayson, D.R. and Darnell, J.E., Jr. (1989) Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes. *Mol Cell Biol,* 9, 1415-25.
61. Overdier, D.G., Porcella, A. and Costa, R.H. (1994) The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Mol Cell Biol,* 14, 2755-66.
62. 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, 683-90.
63. Pierrou, S., Hellqvist, M., Samuelsson, L., Enerback, S. and Carlsson, P. (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. *Embo J,* 13, 5002-12.
64. 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, 5861-9.
65. Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S. and Johnson, R.C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J Mol Biol,* 264, 675-95.
66. Panicker, S.G., Sampath, S., Mandal, A.K., Reddy, A.B., Ahmed, N. and Hasnain, S.E. (2002) Novel mutation in FOXC1 wing region causing Axenfeld-Rieger anomaly. *Invest Ophthalmol Vis Sci,* 43, 3613-6.
67. Komatireddy, S., Chakrabarti, S., Mandal, A.K., Reddy, A.B., Sampath, S., Panicker, S.G. and Balasubramanian, D. (2003) Mutation spectrum of FOXC1 and clinical genetic heterogeneity of Axenfeld-Rieger anomaly in India. *Mol Vis,* 9, 43-8.
68. Melo, F. and Feytmans, E. (1998) Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol,* 277, 1141-52.
69. Saleem, R.A., Murphy, T.C., Liebmann, J.M. and Walter, M.A. (2003) Identification and analysis of a novel mutation in the FOXC1 forkhead domain. *Invest Ophthalmol Vis Sci,* 44, 4608-12.
70. Bravieri, R., Shiyanova, T., Chen, T.H., Overdier, D. and Liao, X. (1997) Different DNA contact schemes are used by two winged helix proteins to recognize a DNA binding sequence. *Nucleic Acids Res,* 25, 2888-96.
71. Zhou, S., Zawel, L., Lengauer, C., Kinzler, K.W. and Vogelstein, B. (1998) Characterization of human FAST-1, a TGF beta and activin signal transducer. *Mol Cell,* 2, 121-7.

72. 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, 351-9.
73. Takiya, S., Gazi, M. and Mach, V. (2003) The DNA binding of insect Fork head factors is strongly influenced by the negative cooperation of neighbouring bases. *Insect Biochem Mol Biol,* 33, 1145-54.
74. Saleem, R.A., Banerjee-Basu, s., Murphy, T.C., Baxevanis, A., and Walter, M.A. (2004) Essential structural and functional determinants within the forkhead domain of FOXC1. *Nucleic Acids Res.*
75. Shiyanova, T. and Liao, X. (1999) The dissociation rate of a winged helix protein-DNA complex is influenced by non-DNA contact residues. *Arch Biochem Biophys,* 362, 356-62.