

Structural Analysis of Disease-Causing Mutations in the P-Subfamily of Forkhead Transcription Factors

Sharmila Banerjee-Basu and Andreas D. Baxevasis*

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland

ABSTRACT Mutations in a number of forkhead transcription factors are associated with the development of inherited diseases in humans. Two closely related genes, *FOXP2* and *FOXP3*, are implicated in two completely different human disorders. A point mutation in the forkhead domain of *FOXP2* (R553H) is responsible for a severe speech and language disorder, while a series of missense mutations distributed over the forkhead domain of *FOXP3* cause a fatal disorder called IPEX, characterized by immune dysregulation, polyendocrinopathy, and enteropathy. Homology model building techniques were used to generate atomic structures of *FOXP2* and *FOXP3*, using the solution structures of the forkhead domain of the adipocyte-transcription factor FREAC-11 and AFX as templates. The impact of these disease-causing missense mutations on the three-dimensional structure, stability, and surface electrostatic charge distribution of the forkhead domains is examined here. The missense mutations R553H in *FOXP2* and R397W in *FOXP3* dramatically alter the electrostatic potentials of the molecular surface of their respective forkhead domains. *Proteins* 2004;54:639–647. © 2004 Wiley-Liss, Inc.

Key words: *FOXP3*; *FOXP2*; IPEX; speech disorders; molecular modeling; surface electrostatic potential

INTRODUCTION

Transcription factors are multidomain proteins, often classified based on the conserved structural motif in their DNA-binding domain. The forkhead domain was first identified as a region of homology between the rat hepatocyte nuclear factor 3 and the *Drosophila melanogaster* transcription factor *Forkhead*.¹ Since then, transcription factors containing the forkhead domain have been identified in a variety of eukaryotic organisms, from yeast to human.² The forkhead domain contains an ~110-residue conserved DNA-binding region that regulates expression of the downstream genes through sequence-specific DNA recognition. Members of the forkhead family of transcription factors play a critical role in diverse developmental processes, including pattern formation, signal transduction, and regulation of tissue-specific cellular differentiation and proliferation.

The three-dimensional (3D) structures of the forkhead domains of HNF-3 (FoxA3), Genesis (FOXDI), AFX, FREAC-11, and ILF-1 have been solved experimentally.^{3–7}

The conserved forkhead domain belongs to the winged helix superfamily of transcription factors. This DNA-binding motif is a variant of the helix-turn-helix (HTH) fold, consisting of 3 α -helices and 2 large loops forming “wing” structures. The major α -helices pack against each other to form the HTH core. This core rests above a small, antiparallel β -sheet from which two loops emerge. The highest degree of sequence conservation in the forkhead domain is found within the helical regions, most noticeably within the third helix. DNA binding occurs mainly through interaction with the third helix; these interactions involve the major groove of the target DNA. The winged regions of the forkhead domain also participate in DNA recognition.³

A number of forkhead transcription factors have been implicated in the development of diverse inherited disorders in human. Mutations in *FOXC1* and *FOXE3* are associated with various developmental defects in the anterior segment of the eye.^{8–10} *FOXL2* mutations cause various eyelid defects, sometimes associated with ovarian failure.^{11,12} *FOXC2* mutations lead to lymphedema of the limbs and double rows of eyelashes.^{13–15} All four of these disorders involve patterning defects in eye development. Disruption in other two-forkhead domain transcription factors lead to immune dysregulation: *FOXP3* mutations lead to IPEX,^{16–18} and *FOXP1* mutations lead to T-cell immunodeficiency.¹⁹ Mutations in *FOXE1* cause thyroid agenesis, cleft palate, and choanal atresia.²⁰ Finally, a point mutation in the *FOXP2* gene cosegregates with a severe speech and language disorder; patients with this disorder have impaired orofacial movements, as well as grammatical and linguistic deficits.^{21,22}

In this article, we present models of the forkhead domains of *FOXP2* and *FOXP3* that have been generated using standard molecular modeling techniques. The homology observed within the forkhead family is relatively high, allowing for reliable structure determination. The models were analyzed based on known experimental information, as well as potential interaction surfaces suggested by conserved exposed residues. The effects of the disease-causing mutations of *FOXP2* and *FOXP3* on the 3D

The Supplementary Materials Referred to in this article can be found at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html>

*Correspondence to: Andreas D. Baxevasis, Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda MD 20892-8002. E-mail: andy@nhgri.nih.gov

Received 1 May 2003; Accepted 25 September 2003

Published online 3 February 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.10621

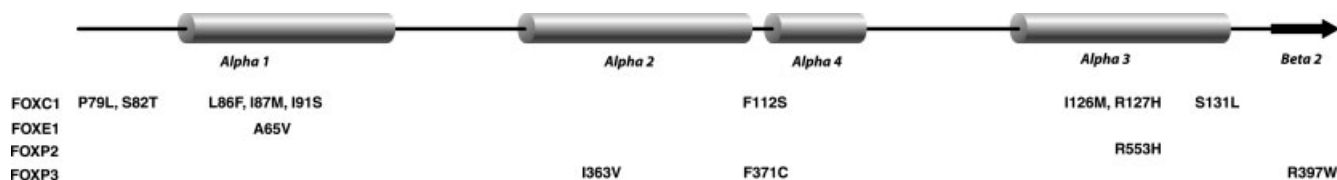


Fig. 1. Missense mutations in the forkhead domains associated with inherited human disorders. The schematic representation includes only the regions where mutations have been documented in the literature.

structures of their respective forkhead domains were examined to gain insight into the underlying causes leading to these disorders.

METHODS

Homology Model Building

Threading experiments were performed by the method of Bryant and Lawrence,²³ with detailed derivations and methodology provided therein. Each query sequence was threaded through the atomic coordinates of the solution structure of the forkhead domain of FREAC-11.³ Three core segments were defined based on the 3D structure. For each possible alignment, individual pairwise residue interactions were determined based on chemical type and distance intervals, lookup tables for which are present in Bryant and Lawrence.²³ Using these values, a conformational energy $\Delta G_{R|M}$, defined as the expected work for substitution of a specific sequence R for a random sequence with the same composition in the context of folding motif M , was then calculated for each alignment. Z -scores ($Z_{R|M}$) and chance occurrence probabilities ($E_{R|M}$) were calculated to compare conformational energies for different alignments. Chance occurrence probabilities give the odds that a random sequence of the same length and amino acid composition would yield a threading energy as low as the query sequence R . Calculations of energies and statistical significance were performed using C and S-PLUS subroutines.²³ Critical interactions are defined as those having a pairwise interaction energy ≤ -1 kcal/mol. All energy scaffold figures were generated using the GRASP software package.²⁴

Modeling of the forkhead domains of FOXP3 and FOXP2 was based on the solved structures of FREAC-11 (pdb|1D5V) and AFX (pdb|1E17). Multiple-sequence alignments were generated using ClustalW,²⁵ followed by manual adjustment in order to better align gaps found outside of the structured regions of the experimentally solved structures. Model building was performed using the MODELLER package,²⁶ as implemented through Insight II (Accelrys, San Diego, CA). The MODELLER program was run in fully automated mode with a high optimization level in order to construct a minimized 3D model of the target sequence by satisfaction of spatial restraints extracted from the template Protein Data Bank (PDB) files. Three models were generated for each of the target sequences. The stereochemistry and geometry of the models were checked using PROCHECK.²⁷ The model with the best parameters, as determined by PROCHECK, was used for building mutant models.

Electrostatic Surface Potential

The electrostatic surface potentials for the molecular models were calculated using Delphi²⁸ and visualized using GRASP.²⁴

Phylogenetic Analysis

Phylogenetic trees for all forkhead P proteins were constructed using algorithms contained within the PHYLIP Phylogeny Inference Package, version 3.5c.²⁹ PROTDIST was used on these sequences to calculate a distance matrix according to the Dayhoff PAM probability model. The computed distances represent the expected fraction of amino acid substitutions between each pair of sequences. The distance matrix was then used to estimate phylogenies using the neighbor-joining (NJ) method.³⁰ Bootstrapping was carried out using SEQBOOT (100 replicates). CONSENSE was used to compute the consensus tree by the majority-rule method. The final unrooted tree diagram was generated using Treetool (http://ftp.sunet.se/pub/science/molecular_biology/unix/treetool/).

RESULTS AND DISCUSSION

The Online Mendelian Inheritance in Man database (OMIM; <http://www.ncbi.nlm.nih.gov/omim>) describes several mutations in forkhead transcription factors that lead to numerous diverse pathological conditions. The disease-causing mutations identified in FOX genes include frame-shift and nonsense mutations that yield truncated protein products lacking the forkhead domain, as well as missense mutations that result in substitution of residues in the forkhead domain. A graphical representation of all documented forkhead domain missense mutations identified in human disorders is shown in Figure 1. It is difficult to explain the biological dysfunction caused by changes as subtle as single-site amino acid residue replacement without directly examining the actual 3D structure of the affected protein. However, the availability of experimentally determined structures of these forkhead domains allows for the assessment of the effect of point mutations on its structure. Previous work from this group involved the characterization of 5 disease-causing missense mutations (S82T, I87M, F112S, I126M, and S131L) in the forkhead domain of FOXC1.⁸ Molecular modeling of FOXC1 predicted that these missense mutations did not destabilize the structure of the forkhead domain; all mutant proteins were correctly expressed and localized to the nucleus, but they introduced subtle changes that interfered with DNA binding or transactivation. We have extended the structural aspects of the study in order to better understand the effect of recently identified missense mutations in FOXC1 (manuscript in preparation), FOXP2, and FOXP3.

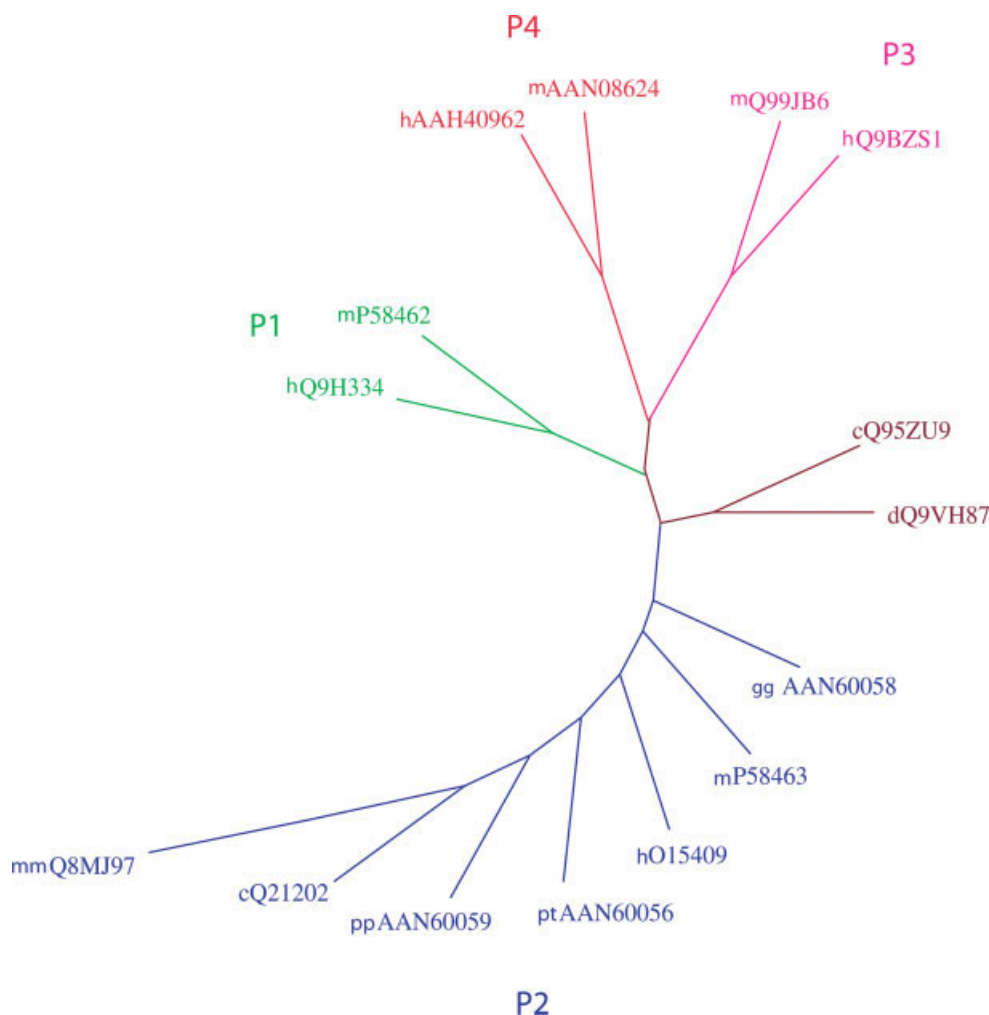


Fig. 2. Phylogenetic relationships between members of the P-type forkhead transcription factors. This unrooted neighbor-joining tree relates the amino acid sequences of the forkhead domains of P-type proteins from human (h), gorilla (gg), chimpanzee (pt), orangutan (pp), monkey (mm), mouse (m), fruit fly (d), and worm (c). Major groupings (P1, P2, P3, and P4) are indicated by color codes. Accession numbers are shown for each sequence; many of these sequences are unnamed but belong to this family.

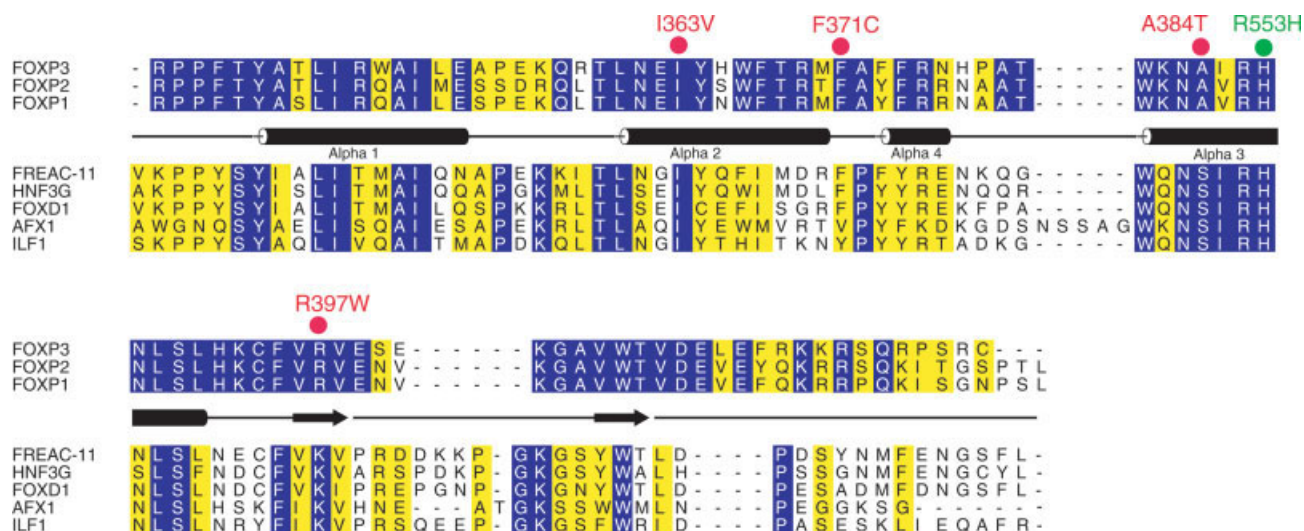


Fig. 3. Multiple-sequence alignment of forkhead domains of selected proteins. The sequences shown in single-letter amino acid codes are from human FOXP1 (Q9H334), FOXP2 (O15409), FOXP3 (Q9BZS1), HNF3 γ (P55318), AFX (P98177), and FREAC-11 (Q99958). Amino acid residues showing absolute identity among these proteins are shown in white against a blue background; those positions with conservative substitutions are shown with a yellow background. The missense mutations of FOXP2 and FOXP3 analyzed in this study are shown as color-coded filled circles above the alignment. The positions of the four α -helices defined in the solution structure of human FREAC-11 are schematically represented in the bar below the alignment. ALSCRIPT was used to format the alignment.³⁴

Forkhead P Proteins

The forkhead family of transcription factors is characterized by a common, 100-residue-long DNA-binding domain known as the forkhead domain. A search of the SMART domain database³¹ (<http://smart.embl-heidelberg.de>) using the sequence of human FOXP3 as the query revealed the presence of an additional zinc finger domain in this protein. The other two known human P-type proteins (FOXP1 and FOXP2) also contain the additional zinc finger domain, located N-terminal to the forkhead domain. A total of 29 proteins having a similar domain organization of zinc finger and forkhead domains were identified. After eliminating fragments and duplicate sequences, 15 unique sequences from 8 species were obtained. A phylogenetic analysis of these 15 sequences identified a novel cluster belonging to FOXP4-type proteins (Fig. 2, shown in red). The forkhead domains of the P-type proteins show a high sequence similarity (80% sequence identity between *Drosophila* and human FOXP1 proteins). Since there is only one P-type protein in *Drosophila* and at least 4 in mammals, this subfamily may have expanded in order to accommodate higher order functions.

Sequence Alignment

The success of homology modeling is critically dependent on a reliable alignment between the target sequences and appropriate structural templates.³² To date, high-resolution 3D structures of the DNA-binding domain for 5 different forkhead transcription factors (FREAC-11,³ AFX,⁴ HNF3- γ ,⁵ ILF-1,⁷ and Genesis⁶) have been experimentally determined. Sequence similarity searches of the P-type proteins showed that their similarity to the experimentally solved structures of the forkhead domain ranges from 39% to 50%. A multiple-sequence alignment of the DNA-binding domains of P-subfamily members, along with forkhead domains for which 3D structures are available, indicates a high degree of sequence conservation within the forkhead family; the conservation is greatest within the structured regions (Fig. 3, blue positions).

The forkhead domains of FOXP3 and FOXP2 have the highest similarity with the known structure of AFX at the sequence level (50% and 47% identity, respectively). Compared to the structures of all the other forkhead domains (FREAC-11,³ HNF3- γ ,⁴ ILF-1,⁷ and Genesis⁶), AFX has an insertion in the loop region preceding helix 3 (Fig. 3). This insertion is also absent from P-type forkhead proteins. In spite of the 5 amino acid residue insertion in the second major loop region, the structure of AFX1 is remarkably similar to the structures of HNF3- γ and FREAC-11, with a root-mean-square deviation (RMSD) for residues in the regular secondary structures of 0.6 and 0.7 Å, respectively.⁴ The forkhead domain of FREAC-11 has 46% sequence identity with that of FOXP3, and without any insertions through all the secondary structural elements except for the loop region preceding the last beta strand (Fig. 3). Considering the sequence and structural similarities, AFX (pdb|1E17) and FREAC-11 (pdb|1D5V) were selected as structural templates to build atomic models of FOXP3 and FOXP2.

Threading Analysis of P-Type Forkhead Domains

Molecular models were generated for the FOXP2 and FOXP3 forkhead domains using threading algorithms developed by Bryant and Lawrence.²³ The 3D structure of the forkhead domain of the adipocyte transcription factor FREAC-11³ was used as a template for the threading analysis (see Methods section). All possible placements of the core segments of the template structure along the query sequence, given the constraints of sequence length, core segment length, and limits of loop length, were considered. Threading contact energies were corrected for sequence composition bias by random shuffling of the aligned residues to generate composition-corrected threading scores ($Z_{R|M}$). To evaluate the statistical significance of computed threading scores, 100 random permutations of the query sequence were generated, and the alignment optimization procedure was then repeated on these now-shuffled sequences. Based on this collection of scores, the probability ($E_{R|M}$) that the threading score for a query sequence would be observed purely by chance was calculated. A summary of the threading results is shown in Table I. Considering the wild-type data for P-subfamily members and FREAC-11, the conformational energies show only a slight difference (~ 1 kcal/mol). More importantly, the probability values ($E_{R|M}$) both meet the statistical criteria of being ≥ 0.05 (5% α -level). As such, it can be concluded that there is a statistically significant match between the sequence of FOXP proteins and the structure of FREAC-11.

Since the threading technique searches for the best alignment of sequence to structure, it is possible to represent the results as a "structural alignment." The structural alignment indicates that there is 46% sequence identity between the FOXP3 and FREAC-11, and 40% sequence identity between FOXP2 and FREAC-11 in their DNA-binding regions, respectively. In addition, there are many positions that are replaced by conservative substitutions within the aligned core segments (Fig. 3, yellow positions). Because of the similarities at the sequence level *and* at the structural level, the model structures of FOXP2 and FOXP3 are deemed to be reliable ones for understanding the significance of point mutations in the forkhead region.

Figure 4 shows the energy scaffolds for the P-type forkhead domains generated in this threading experiment. The energy scaffolds provide a method for visualizing the important intramolecular interactions taking place within a protein. Here, the winged helical bundle is held together by numerous hydrophobic interactions, represented by the thick, magenta-colored cylinders. These interactions also keep the third helix properly positioned for DNA binding. The residues involved in these intermolecular interactions are found primarily between the helical interfaces. The energy scaffolds indicate that the most favorable hydrophobic interactions observed involve Leu 12, Ile 13, and Ile 17 in helix 1; Ile 30, Trp 33, and Phe 34 in helix 2; Trp 48, Ile 52, and Leu 56 in helix 3, and Tyr 77 and Trp 78 in the last β -strand (the numbering is based on the 3D structure of FREAC-11). The amino acid residues participating in these critical pairwise interactions are conserved except for two positions (Trp 33/Phe 34 in P-type pro-

TABLE I. Statistics for Optimal Threads of P-Type Forkhead Domains and Disease-Causing Mutations in FOXP2 and FOXP3

	$\Delta G_{R M}$	$Z_{(s)}$	% Sequence identity	% Structure identity	$P_{(norm)}$
FOXP3 (WT)	121.1	6.2330	81	83	0.0002
R397W	115.6	5.9653	81	76	0.0022
A384T	110.8	5.7215	85	79	0.0206
I363V	109.9	5.5449	84	78	0.0189
F371C	110.2	5.6627	80	74	0.0080
FOXP2 (WT)	105.2	5.7408	77	74	0.0146
R553H	108.1	5.7296	87	84	0.0229
FREAC-11	127.9	6.2536	82	82	0.0036

Sums of contact potentials are expressed as a conformational energy $\Delta G_{R|M}$, the energy associated with nonlocal, nonbonded interactions.²³ Z scores, in standard deviation units, represent the variance from the mean of $\Delta G_{R|M}$. % Sequence identity refers to the fraction of the target sequence aligned. % Structure identity refers to the fraction of the structural domain aligned.

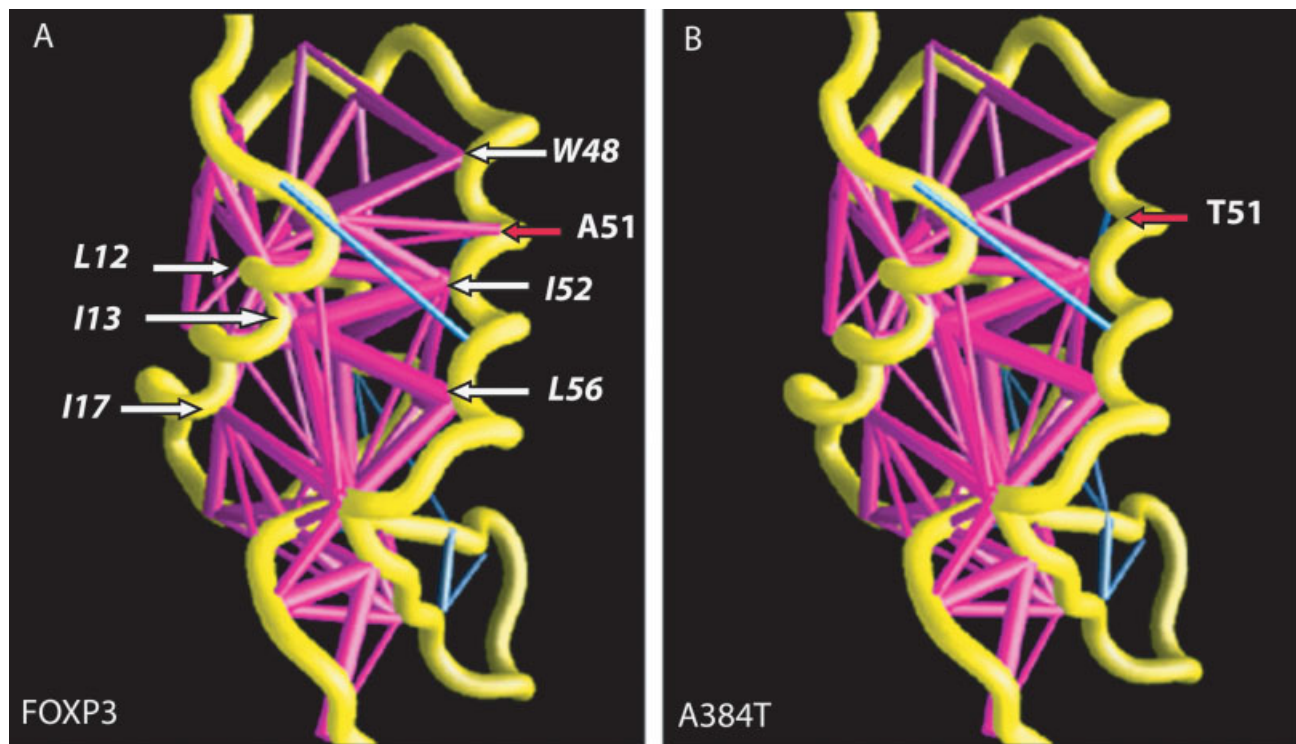


Fig. 4. Molecular models of the P-type forkhead domain. (A) The α -carbon backbone of the molecular model of FOXP3 is depicted as a curving “worm.” Pairwise residue interaction energies between core residues are shown by the width and coloring of the connected α -carbon positions on the protein backbone. Indicated interactions are limited to those with pairwise interaction energies < -1 kcal/mol. Thick, magenta-colored cylinders indicate the most favorable interactions. Intermediate cylinder thicknesses represent interactions with lower pairwise energies. Scaffolds were generated using the graphics program GRASP. (B) Energy scaffolds for the IPEX mutation A384T.

teins are replaced by Phe 33/Ile 34 in FREAC-11). These data, coupled with the information obtained from the multiple-sequence alignments of the entire family, indicate that these key amino acid residues are essential in maintaining the overall fold stability of the forkhead domains.

Structural Model for FOXP2 and FOXP3

Atomic models for the FOXP2 and FOXP3 forkhead domains were generated based on the known 3D structures of FREAC-11 (pdb|1D5V) and AFX (pdb|1E17), that

are most closely related to the P-type forkhead domains. The structural alignment of the P-type proteins generated in the threading experiments was used to make the final alignment for model building. Three different molecular models for FOXP2 and FOXP3 were generated with the restraint-based comparative modeling program MODELLER. Validation of the models was assessed in terms of divergence from the templates and consensus of the alignments using MODELLER target functions (see Supplementary Material). The model with more than 90% of the core in the most favorable region in the Ramachandran plot, as

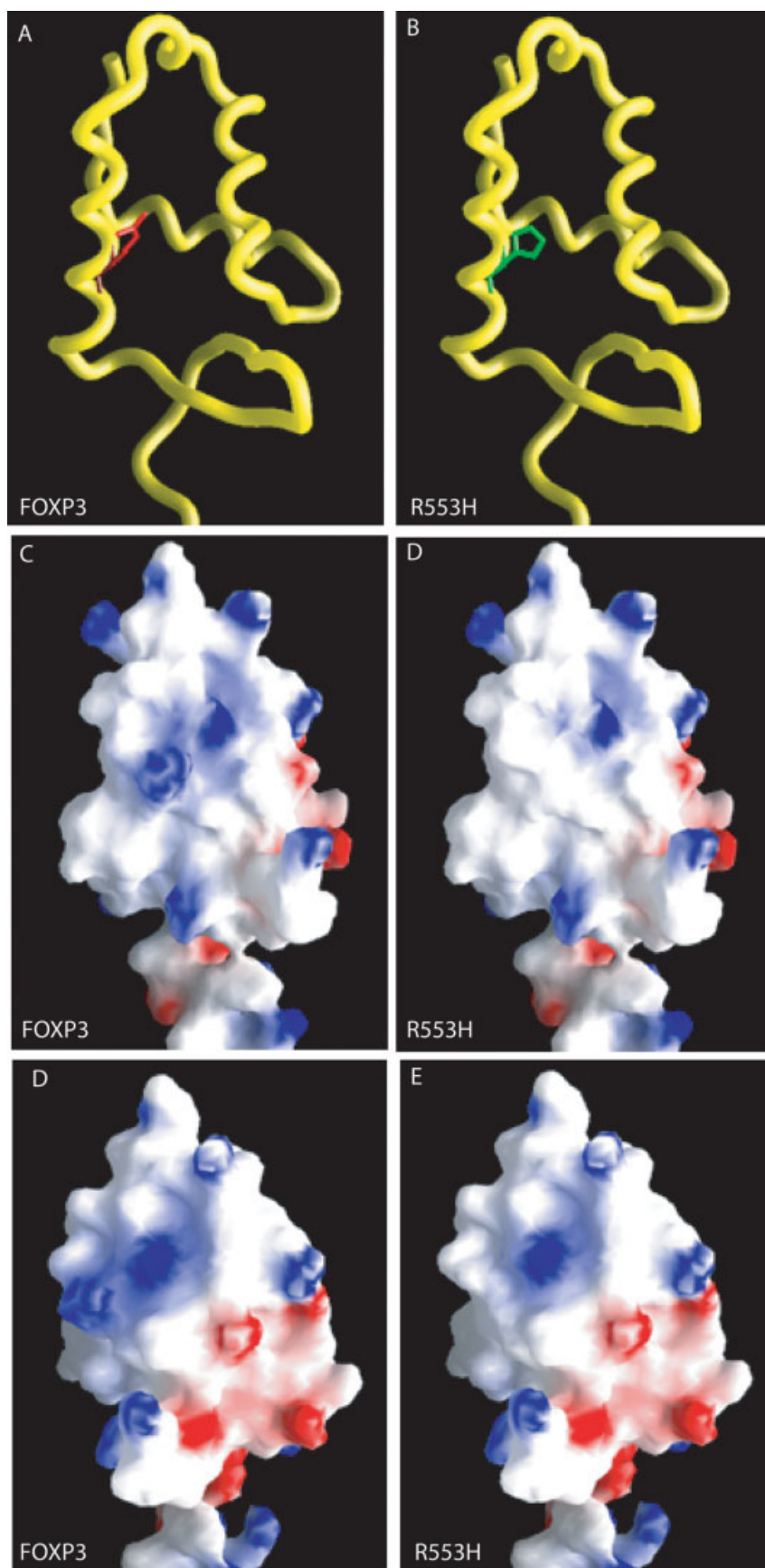


Figure 5.

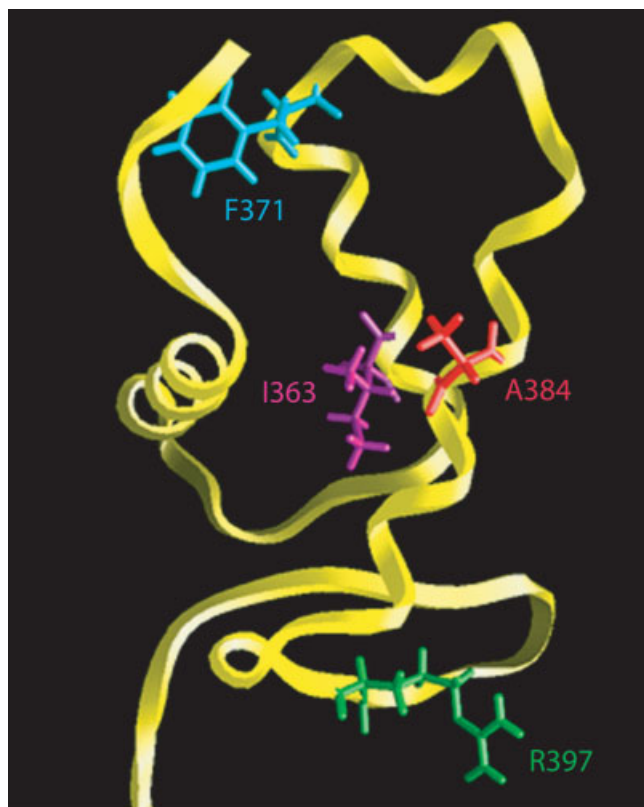


Fig. 6. Molecular models of IPEX mutations. Ribbon model showing the positions of the residues I363V, F371C, A384T, and R397W on the forkhead domain of FOXP3.

assessed by PROCHECK (see Supplementary Material), was used for further analysis.

Detailed examination of the models showed that the maximum variability in the FOXP2 and FOXP3 forkhead structures is observed in the loop region N-terminal to helix 3, as well as in the C-terminal end of helix 3. Both of these regions are critical for DNA recognition. Subtle alterations in the structures are probably responsible for the specificity of the individual forkhead domains.

Structural Analysis of a Speech Mutation

A missense mutation in the forkhead domain of FOXP2 (R553H) is responsible for a speech and language disorder characterized by severe articulation difficulties and deficits in language processing and grammatical skills. The residue involved is one of the invariant amino acid residues among all forkhead transcription factors in all species. In the molecular model of FOXP2, Arg 553 is located on the third helix, pointing away from the hydrophobic core [Fig. 5(panel A)]. Arg 553 is positioned adjacent to a histidine residue that

has been shown to make base-specific contacts in DNA-bound structures of the forkhead domain.⁵ Inspection of the molecular surface of FOXP2 revealed that Arg 553 clusters with Lys 549 and Arg 564 to form a positive surface patch on the DNA-binding face of the forkhead domain. Molecular models of FOXP2 R553H were generated to analyze the structural consequences of this mutation. Replacement of Arg 553 by a histidine residue alters the electrostatic charge distribution on the DNA-binding surface of the FOXP2 forkhead domain [Fig. 5(panels C–F)]. A net reduction in positive charge on the DNA binding surface of the R553H mutant protein may sufficiently alter the determinants required for DNA binding.

Structural Consequences of IPEX Mutations

The FOXP3 forkhead model was used to predict the effect of four missense mutations associated with the lethal developments in IPEX (I363V, F371C, A384T, and R397W). The location of these mutations on the 3D model is shown in Figure 6. All of these missense mutations had threading scores similar to that for wild-type FOXP3 (Table I), indicating that none of these mutations significantly destabilized the overall structure of the forkhead domain. For I363V, replacement of the isoleucine in the second helix by valine did not significantly alter any critical pairwise interactions. F371C, located in the loop region between the second and the third helices of the forkhead domain (Fig. 6), generated threading scores similar to the wild-type FOXP3 (Table I), as would be expected. A favorable pairwise interaction between Tyr 342 and Ala 384 (positions 9 and 51 in the forkhead domain) was eliminated in the A384T mutant [wild-type $\Delta G = -0.95$ kcal/mol, mutated $\Delta G = +0.103$ kcal/mol; Fig. 4(panel B)]. However, the loss of this favorable pairwise interaction, in itself, was not sufficient to destabilize the entire structure of the forkhead domain.

The R397W mutation is located in the conserved second β -strand N-terminal to the first wing [Fig. 7(panel A)]. The corresponding position in the AFX and FREAC-11 templates contain a lysine residue (Fig. 3). Arg 397 contributes to the positively charged surface patch on the DNA-binding face of the FOXP3 forkhead domain [Fig. 7(panel C)]. Substituting the arginine residue with tryptophan, as seen in the R397W mutant, leads to a dramatic alteration in surface charge distribution in FOXP3 [Fig. 7(panel D)]. In particular, note the loss of the positively charged blue area in the mutant model [Fig. 7 (panel D)]. This net reduction in positive charge might interfere with necessary electrostatic interactions between the protein and its target DNA. Additionally, the R397W mutation replaces the longer, more flexible side-chain of arginine with tryptophan's much more rigid, aromatic ring structure [Fig. 7(panels C–D)]. As one would expect, this leads to a significant local distortion of the protein's surface. It can be speculated that the biological processes that involve protein surface recognition might be impaired by these marked changes in both charge distribution and surface geometry.

Fig. 5. Molecular models of R553H mutation. (A and B) Molecular models showing the positions of Arg 553 and His 553 on the forkhead domain of FOXP2. (C) Surface electrostatic charge distribution of FOXP2. (D) Surface electrostatic charge distribution of R553H. (E) Surface model in panel C is rotated 45° clockwise over the Y axis. (F) Surface model in panel D is rotated 45° clockwise over the Y axis. Positive potential is colored in blue; negative potential is colored in red.

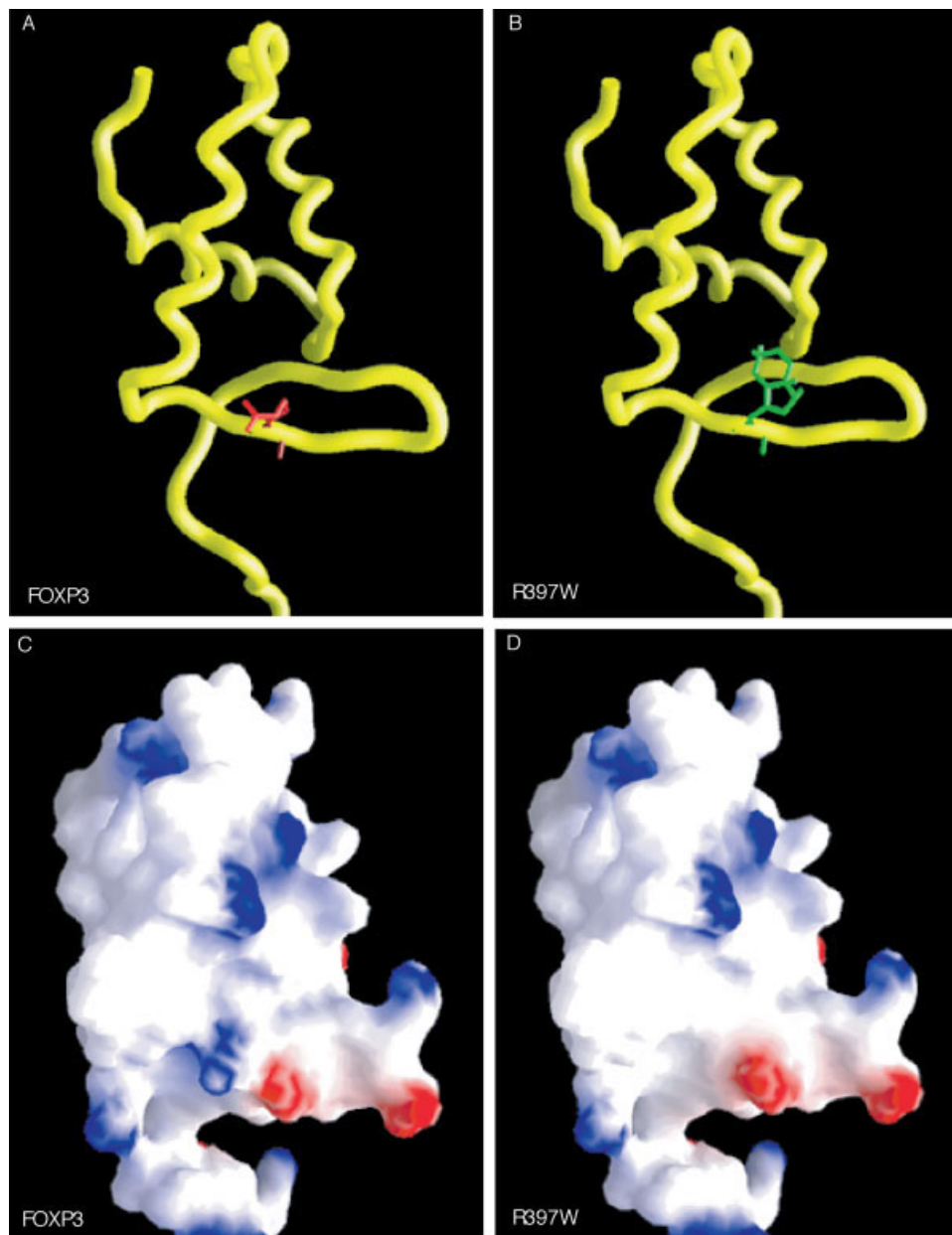


Fig. 7. Molecular models of R397W mutation. (A and B) Molecular models showing the positions of Arg 397 and Trp 397 on the forkhead domain of FOXP3. (C) Surface electrostatic charge distribution of FOXP3. (D) Surface electrostatic charge distribution of R397W. Positive potential is colored in blue; negative potential is colored in red.

Genomic Perspective

The model-building experiments described in this study provide additional insight into the understanding of the structural basis of disease-causing missense mutations in the P subfamily of forkhead transcription factors. These findings provide an important complement to laboratory studies, since they provide a way to better explain observed clinical phenotypes at a structural level. In addition, the identification of critical residues involved in maintaining the stability of P-type forkhead domains will allow for further study of the biochemical properties of this and other forkhead pro-

teins through site-directed mutagenesis and other similar techniques.

In a broader sense, beyond this particular study, the importance of these types of computational studies to genomic science in general and to the understanding of clinical phenotypes in particular cannot be understated. While OMIM and the literature are replete with information on disease-causing mutations, the descriptions of these mutations are usually purely qualitative. It is quite rare for characterizations to go beyond a simple statement that a mutation, deletion, or insertion causes a given phenotype. The availability of genomic sequence data,

along with an ever-increasing number of solved 3D structures, makes it possible to better understand the putative effect of such mutations at the structural level. As described in the "vision document" recently published to mark the end of human genome sequencing,³³ the potential of this type of computational approach—and computational approaches in general—reaches far beyond the study of basic biological processes, finding applicability as "cross-cutting tools" that can have a direct, significant impact on human health.

REFERENCES

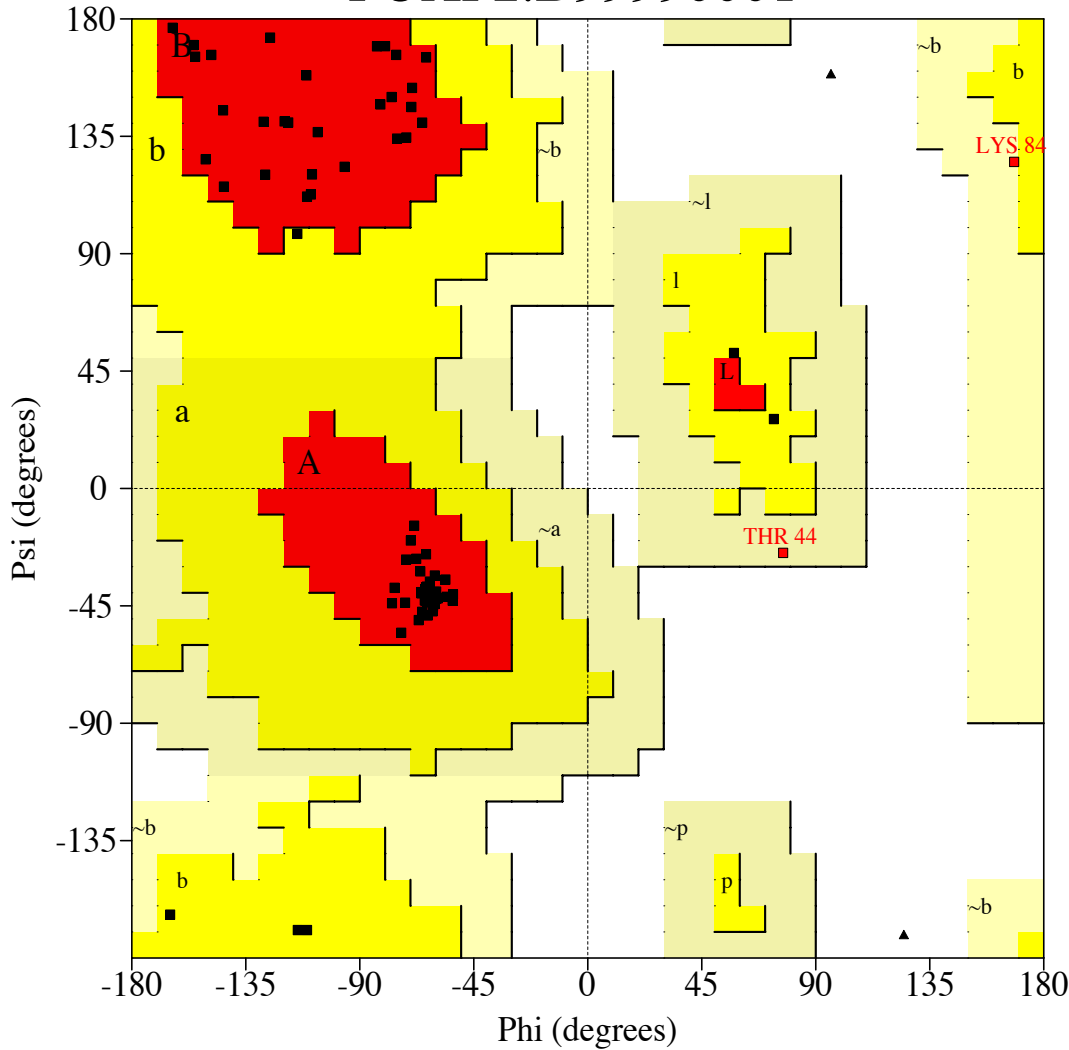
- Weigel D, Jackle H. The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell* 1990;63:455–456.
- Kaufmann E, Knochel W. Five years on the wings of fork head. *Mech Dev* 1996;57:3–20.
- van Dongen MJ, Cederberg A, Carlsson P, Enerback S, Wikstrom M. Solution structure and dynamics of the DNA-binding domain of the adipocyte-transcription factor FREAC-11. *J Mol Biol* 2000;296:351–359.
- Weigelt J, Climent I, Dahlman-Wright K, Wikstrom M. Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). *Biochemistry* 2001;40:5861–5869.
- Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 1993;364:412–420.
- Jin C, Marsden I, Chen X, Liao X. Dynamic DNA contacts observed in the NMR structure of winged helix protein–DNA complex. *J Mol Biol* 1999;289:683–690.
- Liu PP, Chen YC, Li C, Hsieh YH, Chen SW, Chen SH, Jeng WY, Chuang WJ. Solution structure of the DNA-binding domain of interleukin enhancer binding factor 1 (FOXK1a). *Proteins* 2002;49:543–553.
- Saleem RA, Banerjee-Basu S, Berry FB, Baxevanis AD, Walter MA. 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* 2001;68:627–641.
- Lehmann OJ, Ebenezer ND, Jordan T, Fox M, Ocaka L, Payne A, Leroy BP, Clark BJ, Hitchings RA, Povey S, Khaw PT, Bhattacharya SS. Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma. *Am J Hum Genet* 2000;67:1129–1135.
- Ornestad M, Blixt A, Churchill A, Martinsson T, Enerback S, Carlsson P. Foxe3 haploinsufficiency in mice: a model for Peters' anomaly. *Invest Ophthalmol Vis Sci* 2002;43:1350–1357.
- De Baere E, Dixon MJ, Small KW, Jabs EW, Leroy BP, Devriendt K, Gillerot Y, Mortier G, Meire F, Van Maldergem L, Courtens W, Hjalgrim H, Huang S, Liebaers I, Van Regemorter N, Touraine P, Praphanphoj V, Verloes A, Udar N, Yellore V, Chalukya M, Yelchits S, De Paepe A, Kuttann F, Fellous M, Veitia R, Messiaen L. Spectrum of FOXL2 gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype–phenotype correlation. *Hum Mol Genet* 2001;10:1591–1600.
- Fokstuen S, Antonarakis SE, Blouin JL. FOXL2-mutations in blepharophimosis-ptosis-epicanthus inversus syndrome (BPES): challenges for genetic counseling in female patients. *Am J Med Genet* 2003;117A:143–146.
- Traboulsi EI, Al-Khayer K, Matsumoto M, Kimak MA, Crowe S, Wilson SE, Finegold DN, Ferrell RE, Meisler DM. Lymphedema-distichiasis syndrome and FOXC2 gene mutation. *Am J Ophthalmol* 2002;134:592–596.
- Erickson RP. Lymphedema-distichiasis and FOXC2 gene mutations. *Lymphology* 2001;34:1–1.
- Finegold DN, Kimak MA, Lawrence EC, Levinson KL, Cherniske EM, Pober BR, Dunlap JW, Ferrell RE. Truncating mutations in FOXC2 cause multiple lymphedema syndromes. *Hum Mol Genet* 2001;10:1185–1189.
- Tommasini A, Ferrari S, Moratto D, Badolato R, Boniotto M, Pirulli D, Notarangelo LD, Andolina M. X-chromosome inactivation analysis in a female carrier of FOXP3 mutation. *Clin Exp Immunol* 2002;130:127–130.
- Kobayashi I, Shiari R, Yamada M, Kawamura N, Okano M, Yara A, Iguchi A, Ishikawa N, Ariga T, Sakiyama Y, Ochs HD, Kobayashi K. Novel mutations of FOXP3 in two Japanese patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX). *J Med Genet* 2001;38:874–876.
- Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, Whitesell L, Kelly TE, Saulsbury FT, Chance PF, Ochs HD. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet* 2001;27:20–21.
- Frank J, Pignata C, Panteleyev AA, Prowse DM, Baden H, Weiner L, Gaetaniello L, Ahmad W, Pozzi N, Cserhalmi-Friedman PB, Aita VM, Uyttendaele H, Gordon D, Ott J, Brisette JL, Christiano AM. Exposing the human nude phenotype. *Nature* 1999;398:473–474.
- Clifton-Bligh RJ, Wentworth JM, Heinz P, Crisp MS, John R, Lazarus JH, Ludgate M, Chatterjee VK. Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nat Genet* 1998;19:399–401.
- Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 2001;413:519–523.
- Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 2002;418:869–872.
- Bryant SH, Lawrence CE. An empirical energy function for threading protein sequence through the folding motif. *Proteins* 1993;16:92–112.
- Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 1991;11:281–296.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
- Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. *Proteins* 1995;23:318–326.
- Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8:477–486.
- Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, Honig B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J Comput Chem* 2002;23:128–137.
- Felsenstein J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* 1996;266:418–427.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
- Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* 2000;28:231–234.
- Torda AE. Perspectives in protein-fold recognition. *Curr Opin Struct Biol* 1997;7:200–205.
- Collins FS, Green ED, Gutmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003;422:835–847.
- Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* 1993;6:37–40.

Changes in specific pairwise interaction energies resulting from mutating Ala384 to Thr384 in FOXP3

Residue 1	Residue 2	FOXP3 Ala384	A384T Thr384
Thr341	384	0.059	-0.138
Tyr342	384	0.950	-0.103
Leu345	384	0.601	0.434
Ile346	384	0.424	0.413
Leu360	384	0.318	0.308
Ile363	384	0.342	0.333
Tyr364	384	0.193	0.184
Phe367	384	0.305	0.296

Ramachandran Plot

FOXP2.B99990001



Plot statistics		
Residues in most favoured regions [A,B,L]	76	90.5%
Residues in additional allowed regions [a,b,l,p]	6	7.1%
Residues in generously allowed regions [~a,~b,~l,~p]	2	2.4%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	84	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	2	
Number of proline residues	3	
Total number of residues	91	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.



Residues in most favoured regions [A,B,L]	73	91.2%
Residues in additional allowed regions [a,b,l,p]	6	7.5%
Residues in generously allowed regions [~a,~b,~l,~p]	1	1.2%
Residues in disallowed regions	0	0.0%
	----	-----
Number of non-glycine and non-proline residues	80	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	1	
Number of proline residues	5	

Total number of residues	88	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

+-----<<< P R O C E C K S U M M A R Y >>>-----+

	FOXP2.B99990001.pdb	2.0		91 residues
+	Ramachandran plot:	90.5% core	7.1% allow	2.4% gener 0.0% disall
*	All Ramachandrans:	3 labelled residues (out of 89)		
	Chi1-chi2 plots:	0 labelled residues (out of 56)		
	Main-chain params:	6 better	0 inside	0 worse
	Side-chain params:	5 better	0 inside	0 worse
*	Residue properties:	Max.deviation:	2.7	Bad contacts: 2
*		Bond len/angle:	6.5	Morris et al class: 1 1 3
	G-factors	Dihedrals:	0.05	Covalent: -0.23 Overall: -0.05
	M/c bond lengths:	98.7% within limits	1.3% highlighted	
	M/c bond angles:	93.7% within limits	6.3% highlighted	
	Planar groups:	100.0% within limits	0.0% highlighted	

+-----+
 + May be worth investigating further. * Worth investigating further.

+-----<<< P R O C E C K S U M M A R Y >>>-----+

```

FOXP3.B99990002.pdb      2.0                               88 residues
+ Ramachandran plot:    91.2% core      7.5% allow      1.2% gener      0.0% disall
* All Ramachandrans:    4 labelled residues (out of 86)
  Chi1-chi2 plots:      0 labelled residues (out of 59)
  Main-chain params:    6 better        0 inside        0 worse
  Side-chain params:    5 better        0 inside        0 worse
* Residue properties: Max.deviation:      2.5                Bad contacts:      2
*                      Bond len/angle:    6.0      Morris et al class: 1 1 2
G-factors              Dihedrals:    0.07  Covalent:  -0.16      Overall:  -0.02
M/c bond lengths: 98.4% within limits      1.6% highlighted
M/c bond angles:  93.9% within limits      6.1% highlighted
Planar groups:    100.0% within limits      0.0% highlighted

```

+-----<<< P R O C E C K S U M M A R Y >>>-----+

+ May be worth investigating further. * Worth investigating further.