

STRUCTURE NOTE

Structure of the DNA-bound T-box domain of human TBX1, a transcription factor associated with the DiGeorge syndrome

Kamel El Omari,¹ Julie De Mesmaeker,² Dimple Karia,¹ Helen Ginn,¹ Shoumo Bhattacharya,² and Erika J. Mancini^{1*}

¹ Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom

² Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom

Key words: T-box proteins; transcription factors; protein–DNA interaction; crystal structure; DNA binding; patient mutations.

INTRODUCTION

TBX1 (UNIPROT: O43435) is a member of the evolutionary conserved family of T-box transcription factors,¹ which can act as gene enhancer or repressor and are crucial in a wide variety of developmental processes in animals from humans to nematodes.² T-box proteins have a typical molecular weight of 50–78 kDa and are characterized by the presence of a conserved DNA-binding domain, called T-box domain (typically 17–26 kDa). Other domains, such as transcriptional modulating (activating or repressing) domains and nuclear localization motifs are also found in all T-box proteins, although they bear little to no sequence homology.³ The biological importance of these genes is highlighted by the fact that mutations often lead to developmental defects in tissues where they are normally expressed: TBX3 mutations lead to ulnar-mammary syndrome and TBX5 mutations to Holt-Oram syndrome (failure of heart development).⁴ In vertebrates, the T-box proteins can be further classified in five subfamilies: T (Xbra), TBX1, TBX2, TBX6, and TBR subfamilies.⁵ A subset of T-box proteins (TBX1-5, TBX18, and TBX20) are tightly regulated spatially and temporally in developing cardiac structures.³ Although, these T-box proteins can have overlapping expression patterns, each gene has been shown to play an important role in cardiogenesis.⁵

The gene locus for protein TBX1 in humans is on 22q11.2 (chromosome 22). This is the region of a common mutation (named 22q11.2 deletion syndrome or

CATCH22 syndrome) characterized by a deletion of 3 Mb of DNA leading to the phenotype of the DiGeorge syndrome (DGS). DGS affects about one in 4000 births and afflicted babies are born with a variety of different problems, some or all of which may be present to varying degrees. Characteristic signs and symptoms may include thymus and parathyroid defects, facial dysmorphism, pharyngeal insufficiency, congenital heart malformations, and childhood schizophrenia.^{6,7} The cardiac malformations are the major cause of morbidity and mortality in DGS patients.⁸ A high mutation rate in the *TBX1* gene has been observed in DGS patients without the 3 Mb deletion on chromosome 22,⁹ leading to the hypothesis that mutated *TBX1* is the major determinant of DGS.

The T-box domain of T-box proteins binds to a specific DNA consensus sequence TCACACCT, the T-box binding site. Recognition of this site is essential for T-box proteins function but controversy remains in defining the details of the interaction mechanism with DNA. SELEX-experiments have revealed that T-box proteins

Grant sponsor: EU FP7-HEALTH B Program Grant: Cardiogenet; Grant number: 223463; Grant sponsor: BHF and Royal Society.

*Correspondence to: Erika J Mancini, Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX37BN, UK.
E-mail: erika@strubi.ox.ac.uk

Received 8 September 2011; Revised 19 September 2011; Accepted 20 September 2011

Published online 29 September 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.23208

differ in their preference for consensus motifs *in vitro*: TBX5 and TBX20 have been reported to display preferences for a single consensus motif,¹⁰ whereas Xbra and vegT prefer bind paired sites of consensus sequence. These two proteins are selective for different arrangement and spacing of the paired consensus motifs: Xbra would be specific to a head-to-head motif whereas VegT would only bind tail-tail motifs.¹¹ The relevance *in vivo* of palindromic sequence has been however questioned as only half-sites have been identified in natural promoters.

So far, three T-box proteins atomic structures have been determined by X-ray crystallography: Xbra,¹² TBX3,¹³ and TBX5.¹⁴ Xbra and TBX3 were cocrystallized with an *in vitro* selected palindromic DNA, whereas TBX5 was crystallized in both apo form and bound to a single natural T-box binding site. In the work presented here, we have determined the structure of the T-box domain of human TBX1 in complex with a palindromic DNA at 2.6 Å resolution. The structure provides molecular insights into the interaction mechanism of TBX1 with DNA and allows the interpretation of point mutations H194Q and F148Y that have been identified to affect TBX1 function.

MATERIALS AND METHODS

Expression and purification of TBX1 T-box domain

The cDNA coding for an N-terminal histidine tag fused to the human T-box domain (residues 109–297) was synthesized and codon optimized for *E. coli* expression. The restriction sites NdeI and XhoI were respectively added at the 5' and 3' extremities of the fusion gene to enable its insertion in the pET-30a vector (Novagen). The pET-30a-Tbox vector was transformed into Rosetta (DE3) pLysS cells for protein expression and bacterial growth was carried out at 37°C in Luria Broth (LB) supplemented with 50 µg/mL⁻¹ kanamycin until the absorbance at 600 nm reached 0.7. Induction was triggered with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and protein expression was carried over for 16 h at 20°C. The cells were collected by centrifugation and resuspended in 50 mM sodium phosphate pH 7.4, 500 mM NaCl, 10 mM imidazole, and 0.2% (v/v) Tween-20 prior cell disruption by sonication. The lysate was clarified by centrifugation and the supernatant was applied onto a Ni²⁺-charged chelating column equilibrated with lysis buffer without Tween-20. The protein was eluted using a linear imidazole (MP Biomedicals) gradient (10–500 mM). Fractions containing the T-box domain were pooled for a size exclusion purification step using a 16/60 S75 column (GE Healthcare) in a buffer containing 20 mM Tris pH 7.4, 200 mM NaCl and 1 mM DTT. Finally, the T-box domain was concentrated to 20 mg/mL and snap-frozen in liquid nitrogen for long term storage.

Table 1

Data Collection and Refinement Statistics

Data collection	
Space group	$P2_12_12_1$
Unit cell dimensions (Å, °)	$a = 71.4, b = 92.7, c = 101.0,$ $\alpha = \beta = \gamma = 90$
Resolution (Å)	71.4–2.6 (2.64–2.58)
Completeness (%)	98.8 (99.8)
Redundancy	6.4 (6.4)
$\langle I/\sigma \rangle$	11.9 (2.3)
R_{merge} (%)	8.9 (94.7)
Refinement	
Resolution (Å)	68.3–2.6
Number of reflections (measured/unique)	137,384/21,470
$R_{\text{work}}/R_{\text{free}}$ (%)	20.1/22.4
r.m.s.d. bond (Å)	0.008
r.m.s.d. angle (°)	1.1
Mean B-factors/Wilson plot (Å ²)	70.1/79.4
Ramachandran plot (%): favored/allowed/ outliers	98.1/100/0

Values in parentheses are for the highest resolution shell.

Crystallization and structure determination

The synthetic oligonucleotide 5'-AATTTACACCTA GGTGTGAAATT-3' (1 µM scale synthesis, Eurofins MWG Operon) containing the palindromic T-box motif sequence was self-annealed in the presence of 1 mM MgCl₂ in a heat block by slow cooling from 100°C to room temperature. The oligonucleotide was not further purified after annealing. TBX1 and the annealed oligonucleotide were mixed in a 1:1.2 ratio, respectively and the complex was used at 4 mg/mL for crystallization experiments. Initial crystallization trials were performed by the sitting-drop vapor-diffusion method at 292 K using several Hampton and Emerald kits with a Cartesian technologies pipetting robot setting up 100 + 100 nL drops in Greiner 96-well plates.¹⁵ Small crystals appeared within a day in 5% (w/v) Polyethylene Glycol 6000 and 100 mM Citrate pH 5. X-ray diffraction data to 2.6 Å were collected at the microfocus beamline I24 of the Diamond Light Source (Didcot, UK). Diffraction data was indexed, integrated, scaled, and reduced with XIA2.¹⁶ In the highest resolution shell, where the resolution limit cut-off was chosen so that $I/\sigma I > 2$, the redundancy is >6.4 and the R_{merge} is accordingly high ($>94\%$). The space group was $P2_12_12_1$ (unit cell parameters $a = 71.4, b = 92.7, c = 101.0$ Å, and $\alpha = \beta = \gamma = 90^\circ$) with two T-box molecules and one double stranded DNA in the asymmetric unit. Detailed X-ray data collection statistics are given in Table I.

The structure was solved by molecular replacement with MOLREP¹⁷ using a Xbra monomer bound to half site DNA (PDB code 1XBR) as search template. The identified solution was then subjected to rounds of manual rebuilding with the program COOT¹⁸ and NCS restrained refinement with AUTOBUSTER¹⁹ to give final

R_{work} and R_{free} of respectively 20.1% and 22.4%. The final structure was validated with MOLPROBITY²⁰ and deposited with the structure factors in the protein data bank under accession number 4a04.

RESULTS AND DISCUSSION

Overall fold and dimerization interface

The T-box domain of TBX1 is composed of a seven-stranded β -barrel domain core related to an s-type immunoglobulin fold,²¹ and closed by a smaller β -pleated sheet [Fig. 1(A,B)]. As previously defined for Xbra,^{12,13} one sheet is composed of strands A, B, and E, whereas the other comprises strands C, C', F, and G and two other smaller β -sheets (c'cfg and ee'b) close the β -barrel toward the DNA. Helices are found between β -strands and at the C-terminus part of the T-box where helix 3 and 4 are perpendicular to each other and in contact the DNA [Fig. 1(A,B)].

The fold of the T-box domain of TBX1 is very similar to the previously determined T-box structures. TBX1 can be superimposed on Xbra, TBX3, and TBX5 with respective C α rmsd of 0.95, 0.97, and 0.84 Å over 171, 180, and 168 residues [Fig. 1(C)]. Structural comparison of TBX1 with other T-box proteins highlights the insertion of two serines between β -strands c and c', which elongates the loop between them. The major difference lies however in the "dimerization" interface where the length and conformation of the region between β -strands F and G varies greatly [Fig. 1(C)]. The sequence of this region is poorly conserved between T-box proteins and it is one residue longer in TBX1 (ca. 248–254) when compared to TBX5, three compared to TBX3 and seven compared to Xbra [Fig. 1(D)]. This region adopts a 3₁₀ helical conformation in the structure of TBX3 (ca. 239–245) and an extended loop conformation in Xbra (ca. 173–177) with well defined electron density in both cases. In TBX5 this region (ca 190–195) is structurally ill defined with poor electron density visible in either DNA bound or apo structures. In the TBX1 structure presented in this work, the loop between β -strands F and G is only clearly defined in one monomer (b) while no contacts are achieved between monomers [Fig. 1(B)]. It was previously proposed that the order observed in the "dimerization" region in Xbra and TBX3 is induced by their positions on the palindromic DNA, while the absence of such restraints in TBX5 (which is bound to a natural half-palindromic site) allows greater conformational flexibility. It emerges from our structure that in TBX1 the "dimerization" region is mobile even when bound to a DNA bearing the palindromic sequence. We can conclude that this region, where the greatest structural variability amongst T-box domains is observed, is inherently flexible and is probably a protein–protein interaction module specially adapted for each T-box protein.

The arrangement of the two monomers on the palindromic DNA duplex in TBX1 is distinct from the one observed in Xbra or TBX3. It was previously proposed that the relative orientations of the two monomers are also dictated by their underlying position on the palindromic DNA. This appears not be the case as TBX1 was crystallized on the identical palindromic DNA sequence of Xbra, yet it displays a different monomer–monomer orientation.

Understanding the oligomeric state of the T-box proteins is an important question for this family of transcription factor as it directly affects the DNA binding mode. The biological relevance of the dimerization of T-box proteins remains controversial. In absence of DNA, TBX5 crystallizes as a monomer, while in presence of the half-site palindromic DNA it crystallizes as a dimer (interface area 821.5 Å²), arguably a crystallographic artefact dictated by a disulfide bond. It was proposed that Xbra binds a palindromic DNA duplex as dimer¹²; although, the dimer interface is only of 210 Å². The structure of TBX3 bound to a different palindromic DNA duplex suggested that TBX3 binds DNA as two monomers¹³ as their contact areas is only of 185 Å². In the structure of TBX1 presented in this study, no density is observed for the "dimerization" loop in monomer (a) and the interface area is very small (40 Å²). If monomer (b) were to be superimposed on monomer (a) to reconstitute a dimer with two ordered loops, no clashes would be observed but the resulting protein–protein interface would be still too small (140 Å²) to constitute a biologically relevant protein interface. This leads to conclude that like TBX3, TBX1 binds DNA as two monomers and that the dimers are only kept in register by the DNA. A summary of dimer interface areas and solvation energy gains upon dimerization for the available T-box structures is presented in Table II.

A clear picture is now emerging from the structural comparison of several T-box proteins bound to DNA: the dimer interfaces are usually small, their quaternary structure is variable and the sequence underlying the domains responsible for dimerization is not conserved.

DNA binding

DNA is recognized through its major and minor grooves as previously described for the other T-box proteins (Xbra, TBX3, and TBX5). Only few residues are directly involved in DNA binding. In the major groove the only contact is achieved by arginine 137 (R137) with nitrogen N7 and oxygen O2' of the guanine of base pair 5 [Fig. 2(A)]. Nonspecific protein–DNA phosphate backbone interactions are made through hydrogen bonding with polar residues belonging to the sheets bee' and c'cfga.

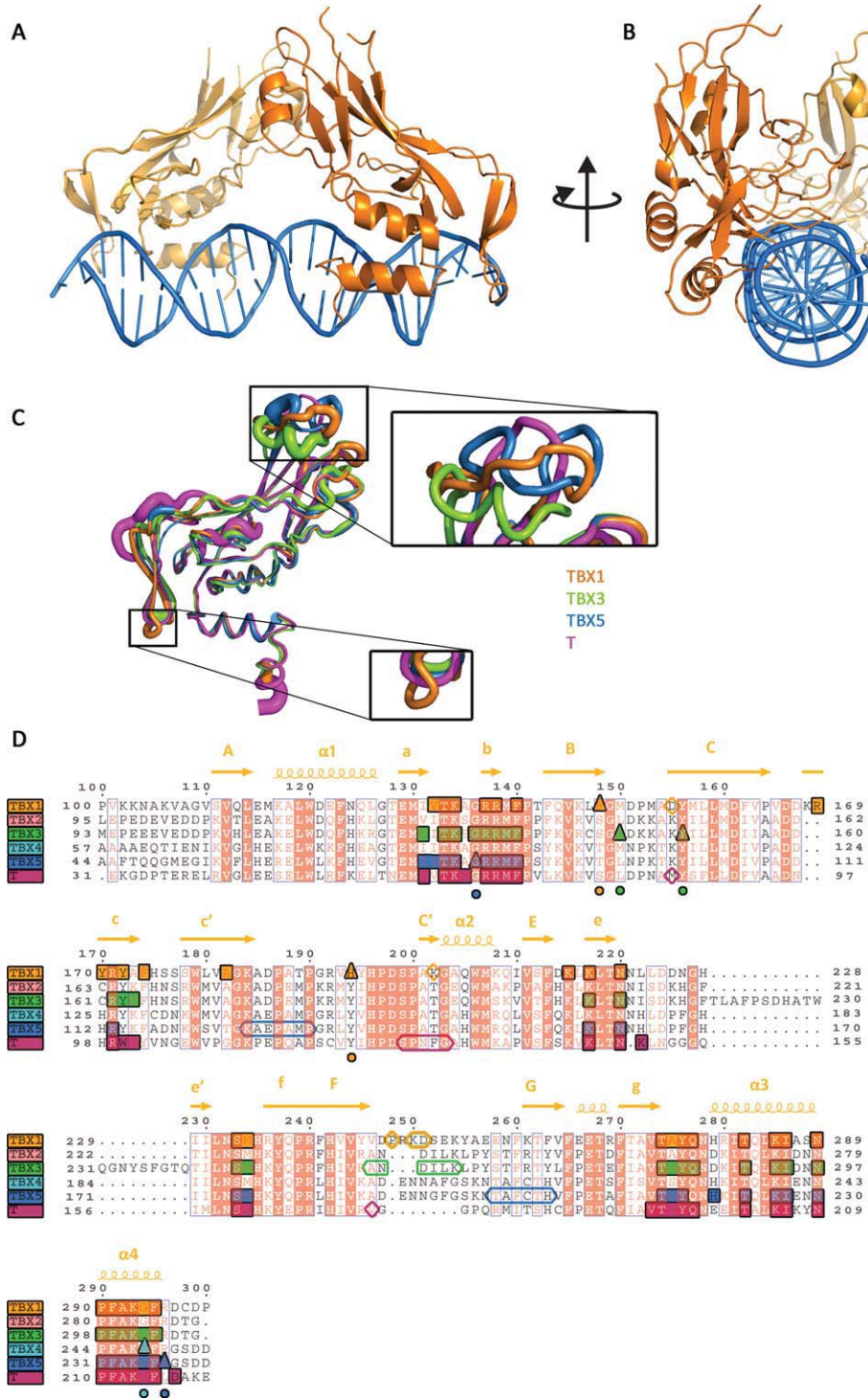


Figure 1

TBX1 T-box structure, structural and sequence alignment with other T-box proteins. (A, B) Cartoon representation of the TBX1-DNA complex in orthogonal views. TBX1 is in yellow (monomer “a” light yellow, monomer “b” dark yellow), DNA is in blue (C) Structural superimposition of TBX1 (dark yellow), TBX3 (green, PDB:1H6F), TBX5 (blue, PDB:2X6U), and T (magenta, PDB:1XBR) in ribbon representation. The thickness of the coil is proportional to the b factor of the residue (D) Sequence alignment of TBX1 and its homologs (TBX1, TBX2, TBX3, TBX4, TBX5, and T). The residue numbers are shown for TBX1, with first and last residues shown for other TBX1 homologs per alignment row. The secondary structure of TBX1 is shown in dark yellow. Residues marked with filled rectangles are those found at the protein-DNA interface. Groups of residues marked with hexagon outlines are found at the dimer interface between T-box protein monomers. Residues marked with triangles show common point mutations implicated in diseases. Marked residues are color-coded for each T-box homolog (TBX1, dark yellow; TBX2, pink; TBX3, green; TBX4, light blue; TBX5, dark blue; T, magenta).

Table II

Summary of Dimer Interface Areas Dimensions and Solvation Energy Gain upon Dimerization

	Dimer interface area ^a (Å ²)	Solvation energy gain ^b (kcal/mol)
TBX1	39.9	1.6
TBX3 (PDB:1HF6)	185.2	-1.6
Xbra (PDB:1XBR)	209.7	-4.2
TBX5 (PDB:2X6V)	821.5	-13.9

^aInterface area in Å², calculated as difference in total accessible surface areas of isolated and interfacing structures divided by two.

^bThe solvation free-energy gain upon formation of the interface, in kcal/M is calculated as difference in total solvation energies of isolated and interfacing structures.

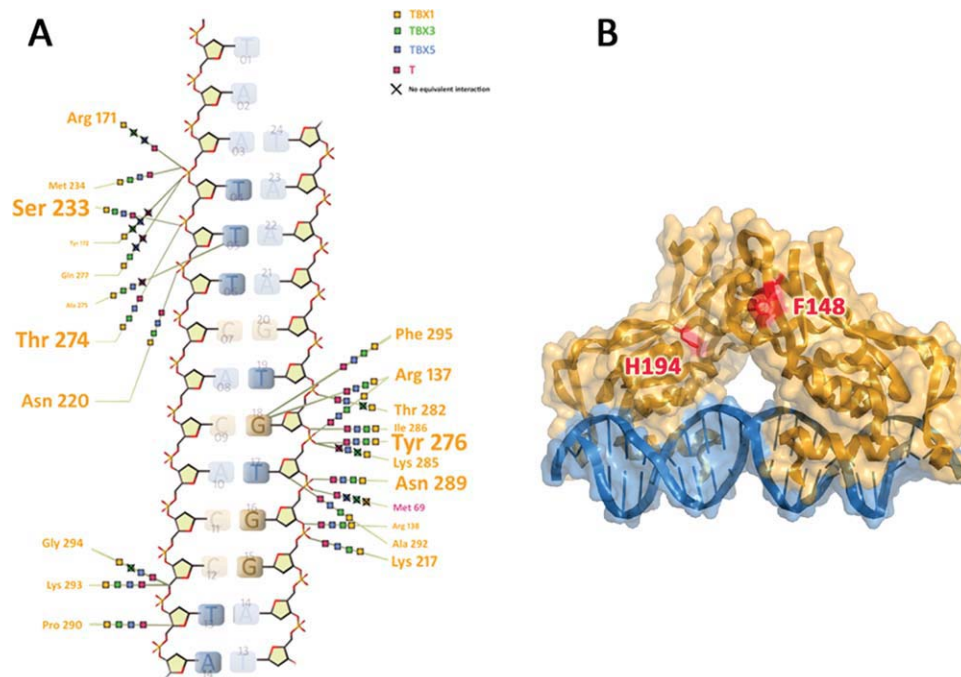
C-terminal helices H3 and H4 which are forming a helix-turn-helix motif recognize the DNA through its minor groove. The contacts involved are hydrogen bonding by tyrosine 276 (Y276) and asparagine 289 (N289) and hydrophobic interactions with the phenylalanine 291 and 295 (F291 and F295). Previous comparison of the DNA-bound and -unbound TBX5 structures showed that helix H4 is either flexible or unwinds in the absence of DNA. Helix H4 was therefore deemed essential for protein-DNA complex formation and the correct orientation of

the transactivation domain downstream the T-box domain.¹⁴ In the TBX5 structure, the DNA used for crystallization corresponds to a single half-site but the binding mode is very similar to the half-site of palindromic DNA used for TBX1, TBX3, and Xbra, confirming the *in vivo* relevance of the binding of T-box domain on the half-sites of palindromic sequences.¹⁴

To allow direct comparison of the DNA binding modes, we cocrystallized TBX1 with the same 24 bp palindromic DNA previously used to cocrystallize Xbra. The comparison shows that TBX1 and Xbra bind DNA in a similar way, when compared to TBX3 bound to a different palindromic DNA sequence. Interestingly, arginine 130 in TBX3 (equivalent to R137 in TBX1) does not contact the guanine base in monomer (b) which could explain the bend observed in the DNA, absent in the Xbra and TBX1 structures.

Analysis of pathogenic TBX1 point mutations

Two point mutations in the T-box domain of TBX1 have been found to disrupt its function: histidine 194 mutation to glutamine (H194Q) and phenylalanine 148 mutation to tyrosine (F148Y).

**Figure 2**

TBX1 T-box-DNA interactions. (A) Map of the DNA showing one half of the palindrome visible in the TBX1 crystal structure, contrasting interactions found in currently determined T-box protein structures. Residues of TBX1 or its homologs are shown interacting using lines with the DNA backbone atoms or specific bases. Each interaction line displays four boxes representing T-box homologs (TBX1, dark yellow; TBX3, green; TBX5, dark blue; and T, magenta), which are crossed out if there is no equivalent interaction found in that homolog. Residue names displayed are for TBX1 residues where possible except for Met69 (Xbra protein only). The size of the residue name text correlates with the proportion of the side chain surface involved in the interaction with DNA. (B) Cartoon representation of the TBX1-DNA complex highlighting the location of mutations H194Q and F148Y.

The H194Q mutation has been found in patients presenting Shprintzen syndrome (velocardio facial). It is clear from our structure that H194, which sits on an extended strand, is neither involved in DNA binding nor dimer formation [Fig. 2(B)]. In previous studies, the histidine to glutamine mutation was thought to create a novel hydrogen bond to the carbonyl oxygen of glycine 227 (G227), which would in turn stabilize the fold of the protein.²² It is unlikely that the mutation plays a role in the formation of a new intra-molecular hydrogen bond as histidine and glutamine residues have approximately the same length and both have a nitrogen atom in their side chain to contact the G227 carbonyl. On the other hand, following the mutation to glutamine the solvent accessibility of the residue is predicted to increase going from buried to partially accessible; the mutation is therefore likely to stabilize the fold of the protein.

The F148Y mutation was found in patients diagnosed with conotruncal anomaly face syndrome,²³ one of the various phenotypes of DGS. Since, F148 is located in the conserved T-box domain this mutation was thought to impair either DNA binding or dimerization. Analysis of our structure reveals that F148 is not located at the DNA or dimerization interface, but is in fact exposed at the surface of the protein. As exposed hydrophobic residues are usually very good candidates for protein interaction we propose that this amino-acid is involved in the interaction with a different protein partner.

CONCLUSIONS

In summary we report the crystal structure of the T-box domain of TBX1, a key transcription factor in cardiogenesis whose mutations can cause dramatic physical abnormality. The structure of TBX1 bound to a palindromic DNA reveals that its T-box domain binds DNA as two distinct monomers whose relative orientation is unique when compared to other T-box proteins and whose contacts with DNA are dictated by the underlying sequence. Finally, analysis of the TBX1 structure has allowed a better mechanistic understanding of two clinically relevant mutations.

REFERENCES

1. Showell C, Binder O, Conlon FL. T-box genes in early embryogenesis. *Dev Dyn* 2004;229:201–218.
2. Smith J. T-box genes: what they do and how they do it. *Trends Genet* 1999;15:154–158.
3. Plageman TF, Jr, Yutzey KE. T-box genes and heart development: putting the “T” in heart. *Dev Dyn* 2005;232:11–20.
4. Wilson V, Conlon FL. The T-box family. *Genome Biol* 2002;3:REVIEWS3008.

5. Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE. T-box genes in vertebrate development. *Annu Rev Genet* 2005;39:219–239.
6. Botto LD, May K, Fernhoff PM, Correa A, Coleman K, Rasmussen SA, Merritt RK, O’Leary LA, Wong LY, Elixson EM, Mahle WT, Campbell RM. A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population. *Pediatrics* 2003;112(Pt 1):101–107.
7. Scambler PJ. The 22q11 deletion syndromes. *Hum Mol Genet* 2000;9:2421–2426.
8. Ryan AK, Rosenfeld MG. POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev* 1997;11:1207–1225.
9. Baldini A. Dissecting contiguous gene defects: TBX1. *Curr Opin Genet Dev* 2005;15:279–284.
10. Macindoe I, Glockner L, Vukasin P, Stennard FA, Costa MW, Harvey RP, Mackay JP, Sunde M. Conformational stability and DNA binding specificity of the cardiac T-box transcription factor Tbx20. *J Mol Biol* 2009;389:606–618.
11. Conlon FL, Fairclough L, Price BM, Casey ES, Smith JC. Determinants of T box protein specificity. *Development* 2001;128:3749–3758.
12. Muller CW, Herrmann BG. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 1997;389:884–888.
13. Coll M, Seidman JG, Muller CW. Structure of the DNA-bound T-box domain of human TBX3, a transcription factor responsible for ulnar-mammary syndrome. *Structure* 2002;10:343–356.
14. Stirnimann CU, Ptchelkine D, Grimm C, Muller CW. Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. *J Mol Biol* 2010;400:71–81.
15. Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC, Berrow NS, Brown J, Berry IM, Stewart-Jones GB, Grimes JM, Stammers DK, Esnouf RM, Jones EY, Owens RJ, Stuart DI, Harlos K. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* 2005;61(Pt 6):651–657.
16. Winter G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 2010;43:186–190.
17. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Cryst* 1997;30:1022–1025.
18. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60(12 Pt 1):2126–2132.
19. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Smart OS, Vonnrhein C, Womack T. Global Phasing Ltd, Cambridge. BUSTER-TNT 2.5.1 and autoBUSTER 1.3.1. Cambridge: Global Phasing Ltd; 2008.
20. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, III, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007;35:W375–W383 (Web Server issue).
21. Bork P, Holm L, Sander C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* 1994;242:309–320.
22. Zweier C, Sticht H, Aydin-Yaylagul I, Campbell CE, Rauch A. Human TBX1 missense mutations cause gain of function resulting in the same phenotype as 22q11.2 deletions. *Am J Hum Genet* 2007;80:510–517.
23. Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S, Ichida F, Joo K, Kimura M, Imamura S, Kamatani N, Momma K, Takao A, Nakazawa M, Shimizu N, Matsuoka R. Role of TBX1 in human del22q11.2 syndrome. *Lancet* 2003;362:1366–1373.