



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 07:57 am GMT

PDB ID : 4A04  
Title : Structure of the DNA-bound T-box domain of human TBX1, a transcription factor associated with the DiGeorge syndrome.  
Authors : El Omari, K.; De Mesmaeker, J.; Karia, D.; Ginn, H.; Bhattacharya, S.; Mancini, E.J.  
Deposited on : 2011-09-07  
Resolution : 2.58 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

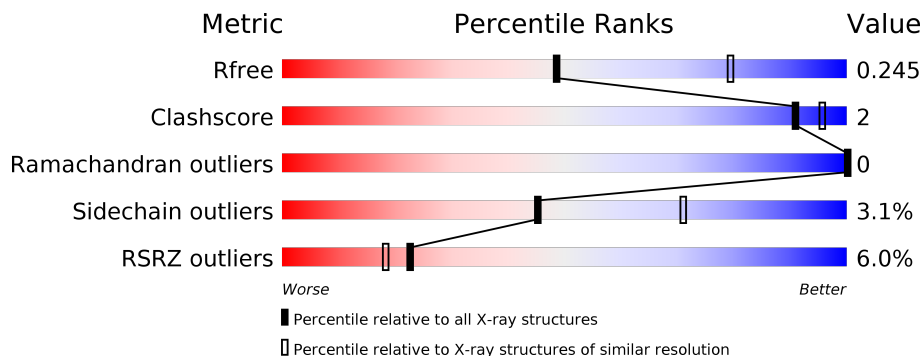
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.58 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	2899 (2.60-2.56)
Clashscore	112137	3268 (2.60-2.56)
Ramachandran outliers	110173	3218 (2.60-2.56)
Sidechain outliers	110143	3218 (2.60-2.56)
RSRZ outliers	101464	2907 (2.60-2.56)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	203	<div> <div>8%</div> <div> <div></div> <div>81%</div> <div>7%</div> <div>11%</div> </div> </div>
1	B	203	<div> <div>4%</div> <div> <div></div> <div>84%</div> <div>8%</div> <div>6%</div> </div> </div>
2	D	24	<div> <div>29%</div> <div>63%</div> <div></div> </div>
2	E	24	<div> <div>4%</div> <div> <div></div> <div>33%</div> <div>67%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4062 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called T-BOX TRANSCRIPTION FACTOR TBX1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	181	Total	C	N	O	S	0	0	0
			1484	956	259	260	9			
1	B	190	Total	C	N	O	S	0	0	0
			1555	999	270	277	9			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	MET	-	EXPRESSION TAG	UNP O43435
A	96	HIS	-	EXPRESSION TAG	UNP O43435
A	97	HIS	-	EXPRESSION TAG	UNP O43435
A	98	HIS	-	EXPRESSION TAG	UNP O43435
A	99	HIS	-	EXPRESSION TAG	UNP O43435
A	100	HIS	-	EXPRESSION TAG	UNP O43435
A	101	HIS	-	EXPRESSION TAG	UNP O43435
A	102	GLU	-	EXPRESSION TAG	UNP O43435
A	103	ASN	-	EXPRESSION TAG	UNP O43435
A	104	LEU	-	EXPRESSION TAG	UNP O43435
A	105	TYR	-	EXPRESSION TAG	UNP O43435
A	106	PHE	-	EXPRESSION TAG	UNP O43435
A	107	GLN	-	EXPRESSION TAG	UNP O43435
A	108	GLY	-	EXPRESSION TAG	UNP O43435
B	95	MET	-	EXPRESSION TAG	UNP O43435
B	96	HIS	-	EXPRESSION TAG	UNP O43435
B	97	HIS	-	EXPRESSION TAG	UNP O43435
B	98	HIS	-	EXPRESSION TAG	UNP O43435
B	99	HIS	-	EXPRESSION TAG	UNP O43435
B	100	HIS	-	EXPRESSION TAG	UNP O43435
B	101	HIS	-	EXPRESSION TAG	UNP O43435
B	102	GLU	-	EXPRESSION TAG	UNP O43435
B	103	ASN	-	EXPRESSION TAG	UNP O43435
B	104	LEU	-	EXPRESSION TAG	UNP O43435
B	105	TYR	-	EXPRESSION TAG	UNP O43435

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	106	PHE	-	EXPRESSION TAG	UNP O43435
B	107	GLN	-	EXPRESSION TAG	UNP O43435
B	108	GLY	-	EXPRESSION TAG	UNP O43435

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	23	Total	C	N	O	P	0	0	0
			469	226	86	135	22			
2	E	24	Total	C	N	O	P	0	0	0
			488	236	88	141	23			

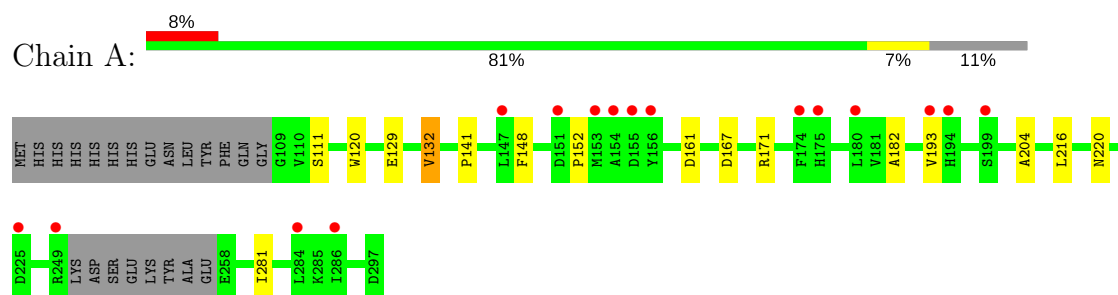
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	20	Total	O	0	0
			20	20		
3	B	24	Total	O	0	0
			24	24		
3	D	15	Total	O	0	0
			15	15		
3	E	7	Total	O	0	0
			7	7		

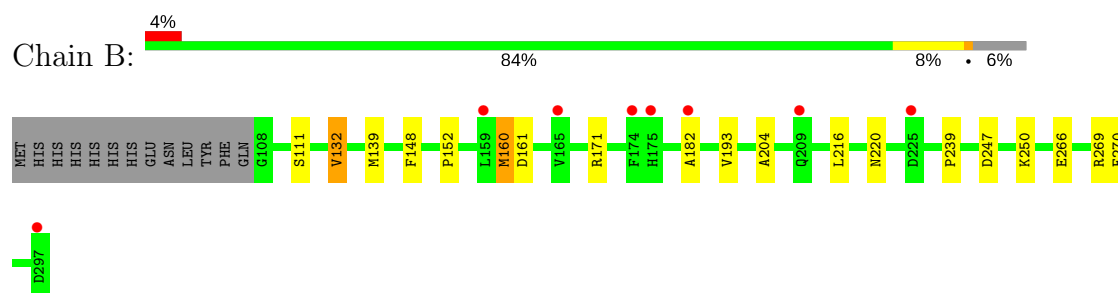
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: T-BOX TRANSCRIPTION FACTOR TBX1



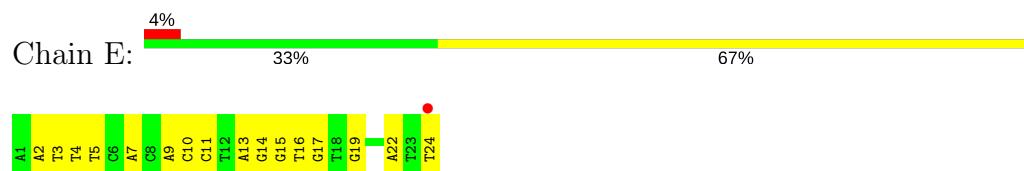
#### • Molecule 1: T-BOX TRANSCRIPTION FACTOR TBX1



#### • Molecule 2: DNA



#### • Molecule 2: DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.42Å 92.74Å 101.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.32 – 2.58 68.32 – 2.58	Depositor EDS
% Data completeness (in resolution range)	98.8 (68.32-2.58) 98.8 (68.32-2.58)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 2.58Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, $R_{free}$	0.206 , 0.227 0.216 , 0.245	Depositor DCC
$R_{free}$ test set	1100 reflections (5.40%)	DCC
Wilson B-factor (Å <sup>2</sup> )	61.3	Xtriage
Anisotropy	0.596	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 41.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4062	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.20% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.38	0/1526	0.59	0/2062
1	B	0.39	0/1599	0.63	0/2160
2	D	0.90	0/526	1.69	16/810 (2.0%)
2	E	0.89	0/547	1.70	16/843 (1.9%)
All	All	0.56	0/4198	1.04	32/5875 (0.5%)

There are no bond length outliers.

All (32) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	13	DA	P-O3'-C3'	10.08	131.80	119.70
2	E	22	DA	P-O3'-C3'	9.92	131.61	119.70
2	D	22	DA	P-O3'-C3'	9.78	131.44	119.70
2	E	13	DA	P-O3'-C3'	9.16	130.70	119.70
2	E	14	DG	O4'-C1'-N9	8.63	114.05	108.00
2	D	14	DG	O4'-C1'-N9	8.63	114.04	108.00
2	D	19	DG	P-O3'-C3'	8.15	129.48	119.70
2	E	19	DG	P-O3'-C3'	7.57	128.78	119.70
2	E	11	DC	O4'-C1'-N1	7.47	113.23	108.00
2	D	17	DG	P-O3'-C3'	6.95	128.04	119.70
2	D	2	DA	P-O3'-C3'	6.79	127.85	119.70
2	D	3	DT	P-O3'-C3'	6.79	127.85	119.70
2	D	11	DC	O4'-C1'-N1	6.66	112.66	108.00
2	E	17	DG	P-O3'-C3'	6.48	127.48	119.70
2	D	4	DT	P-O3'-C3'	6.41	127.39	119.70
2	E	2	DA	P-O3'-C3'	6.30	127.27	119.70
2	E	4	DT	P-O3'-C3'	6.24	127.19	119.70
2	E	10	DC	O4'-C1'-N1	6.23	112.36	108.00
2	E	9	DA	P-O3'-C3'	6.14	127.07	119.70
2	D	9	DA	P-O3'-C3'	5.99	126.89	119.70
2	D	7	DA	P-O3'-C3'	5.98	126.87	119.70
2	E	5	DT	P-O3'-C3'	5.88	126.76	119.70
2	E	7	DA	P-O3'-C3'	5.76	126.61	119.70

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	10	DC	O4'-C1'-N1	5.65	111.95	108.00
2	D	15	DG	P-O3'-C3'	5.60	126.42	119.70
2	E	3	DT	P-O3'-C3'	5.54	126.35	119.70
2	D	5	DT	P-O3'-C3'	5.41	126.20	119.70
2	E	15	DG	P-O3'-C3'	5.37	126.14	119.70
2	E	16	DT	O4'-C1'-N1	5.33	111.73	108.00
2	D	16	DT	P-O3'-C3'	5.18	125.92	119.70
2	D	14	DG	P-O3'-C3'	5.14	125.87	119.70
2	E	24	DT	C4-C5-C7	5.05	122.03	119.00

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1484	0	1453	6	0
1	B	1555	0	1518	8	0
2	D	469	0	262	1	0
2	E	488	0	271	0	0
3	A	20	0	0	0	0
3	B	24	0	0	0	0
3	D	15	0	0	0	0
3	E	7	0	0	0	0
All	All	4062	0	3504	15	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

All (15) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:171:ARG:HB2	1:B:182:ALA:HB2	1.87	0.57
1:B:111:SER:HB2	1:B:148:PHE:HB2	1.88	0.55
1:A:111:SER:HB2	1:A:148:PHE:HB2	1.88	0.55

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:171:ARG:HB2	1:A:182:ALA:HB2	1.89	0.54
1:A:132:VAL:CG1	1:A:220:ASN:HB3	2.39	0.53
1:B:160:MET:HE2	1:B:216:LEU:HD13	1.98	0.46
1:A:129:GLU:HB2	1:A:281:ILE:HG21	1.98	0.46
1:B:152:PRO:HB3	1:B:204:ALA:HB2	1.99	0.45
1:A:152:PRO:HB3	1:A:204:ALA:HB2	1.99	0.44
1:B:132:VAL:HG13	1:B:220:ASN:HB3	2.00	0.43
1:B:247:ASP:HB3	1:B:250:LYS:HD2	2.01	0.42
1:B:266:GLU:O	1:B:269:ARG:HG2	2.20	0.41
1:B:239:PRO:HG3	1:B:270:PHE:HE2	1.85	0.41
2:D:1:DA:H2'	2:D:2:DA:C8	2.55	0.41
1:A:120:TRP:CE2	1:A:141:PRO:HG2	2.56	0.41

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	177/203 (87%)	173 (98%)	4 (2%)	0	100	100
1	B	188/203 (93%)	184 (98%)	4 (2%)	0	100	100
All	All	365/406 (90%)	357 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was

analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	159/179 (89%)	154 (97%)	5 (3%)	45	71
1	B	166/179 (93%)	161 (97%)	5 (3%)	46	72
All	All	325/358 (91%)	315 (97%)	10 (3%)	45	71

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	132	VAL
1	A	161	ASP
1	A	167	ASP
1	A	193	VAL
1	A	216	LEU
1	B	132	VAL
1	B	139	MET
1	B	160	MET
1	B	161	ASP
1	B	193	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	228	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

There are no ligands in this entry.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	181/203 (89%)	0.56	16 (8%) 11 8	42, 67, 97, 163	0
1	B	190/203 (93%)	0.42	8 (4%) 37 31	42, 66, 101, 129	0
2	D	23/24 (95%)	-0.45	0 100 100	51, 65, 77, 83	0
2	E	24/24 (100%)	-0.29	1 (4%) 37 31	57, 69, 85, 112	0
All	All	418/454 (92%)	0.39	25 (5%) 23 18	42, 67, 100, 163	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	297	ASP	6.4
1	A	249	ARG	4.4
1	A	153	MET	3.2
1	B	182	ALA	3.1
2	E	24	DT	3.0
1	B	175	HIS	2.9
1	B	225	ASP	2.9
1	A	151	ASP	2.9
1	A	199	SER	2.8
1	A	155	ASP	2.6
1	A	193	VAL	2.6
1	B	165	VAL	2.5
1	A	154	ALA	2.5
1	B	174	PHE	2.4
1	A	194	HIS	2.4
1	B	209	GLN	2.3
1	A	156	TYR	2.3
1	A	174	PHE	2.3
1	A	175	HIS	2.2
1	A	180	LEU	2.1
1	B	159	LEU	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	286	ILE	2.1
1	A	147	LEU	2.0
1	A	284	LEU	2.0
1	A	225	ASP	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.