



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 14, 2017 – 05:50 AM EST

PDB ID : 5T1J  
Title : Crystal Structure of the Tbox DNA binding domain of the transcription factor T-bet  
Authors : Liu, C.F.; Brandt, G.S.; Hoang, Q.; Hwang, E.S.; Naumova, N.; Lazarevic, V.; Dekker, J.; Glimcher, L.H.; Ringe, D.; Petsko, G.A.  
Deposited on : unknown  
Resolution : 2.95 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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	:	rb-20030345
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)	:	rb-20030345

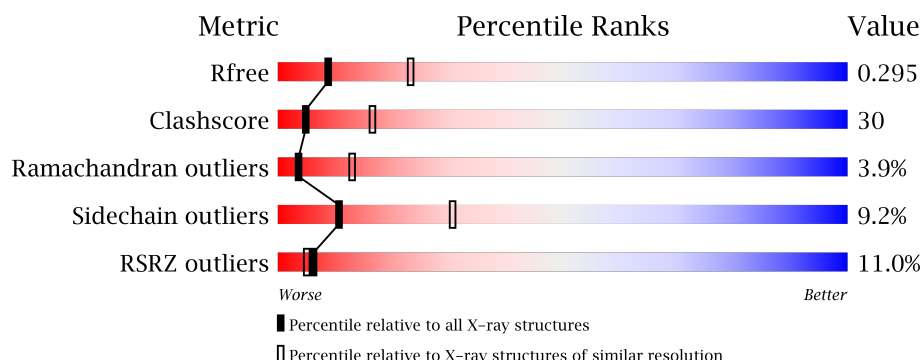
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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	2289 (2.98-2.90)
Clashscore	112137	2543 (2.98-2.90)
Ramachandran outliers	110173	2475 (2.98-2.90)
Sidechain outliers	110143	2477 (2.98-2.90)
RSRZ outliers	101464	2301 (2.98-2.90)

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	204	<div> <div>10%</div> <div> <div>43%</div> <div>43%</div> <div>8%</div> <div>6%</div> </div> </div>
1	B	204	<div> <div>13%</div> <div> <div>37%</div> <div>49%</div> <div>8%</div> <div>6%</div> </div> </div>
2	C	24	<div> <div>4%</div> <div> <div>46%</div> <div>46%</div> <div>8%</div> </div> </div>
2	D	24	<div> <div>4%</div> <div> <div>42%</div> <div>54%</div> <div>.</div> </div> </div>
2	E	24	<div> <div>8%</div> <div> <div>46%</div> <div>38%</div> <div>17%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	24	<div><div></div><div>8%</div><div>25%</div><div>67%</div><div>8%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5062 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 TBX21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	192	Total	C	N	O	S	0	0	0
			1553	983	282	280	8			
1	B	192	Total	C	N	O	S	0	0	0
			1553	983	282	280	8			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	initiating methionine	UNP Q9JKD8
A	125	ARG	-	expression tag	UNP Q9JKD8
A	126	GLY	-	expression tag	UNP Q9JKD8
A	127	SER	-	expression tag	UNP Q9JKD8
A	128	HIS	-	expression tag	UNP Q9JKD8
A	129	HIS	-	expression tag	UNP Q9JKD8
A	130	HIS	-	expression tag	UNP Q9JKD8
A	131	HIS	-	expression tag	UNP Q9JKD8
A	132	HIS	-	expression tag	UNP Q9JKD8
A	133	HIS	-	expression tag	UNP Q9JKD8
A	134	GLY	-	expression tag	UNP Q9JKD8
A	135	SER	-	expression tag	UNP Q9JKD8
B	124	MET	-	initiating methionine	UNP Q9JKD8
B	125	ARG	-	expression tag	UNP Q9JKD8
B	126	GLY	-	expression tag	UNP Q9JKD8
B	127	SER	-	expression tag	UNP Q9JKD8
B	128	HIS	-	expression tag	UNP Q9JKD8
B	129	HIS	-	expression tag	UNP Q9JKD8
B	130	HIS	-	expression tag	UNP Q9JKD8
B	131	HIS	-	expression tag	UNP Q9JKD8
B	132	HIS	-	expression tag	UNP Q9JKD8
B	133	HIS	-	expression tag	UNP Q9JKD8
B	134	GLY	-	expression tag	UNP Q9JKD8
B	135	SER	-	expression tag	UNP Q9JKD8

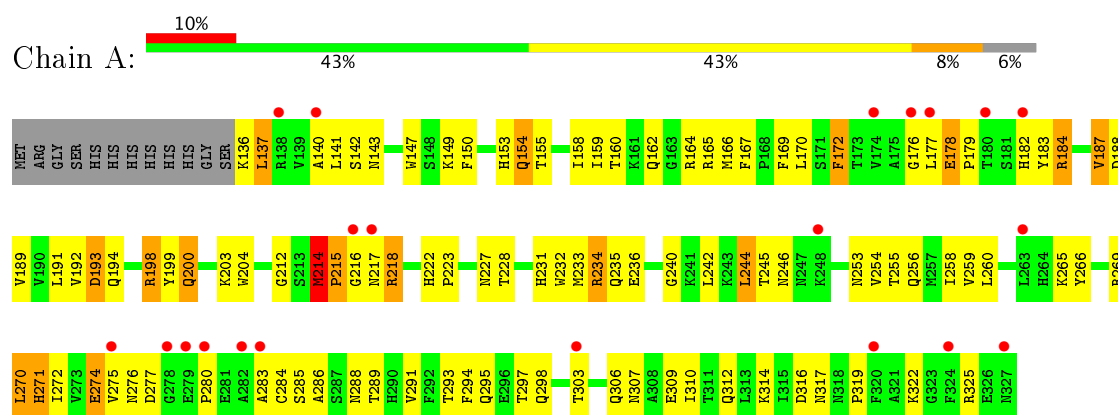
- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	24	Total 489	C 236	N 88	O 142	P 23	0	0	0
2	D	24	Total 489	C 236	N 88	O 142	P 23	0	0	0
2	E	24	Total 489	C 236	N 88	O 142	P 23	0	0	0
2	F	24	Total 489	C 236	N 88	O 142	P 23	0	0	0

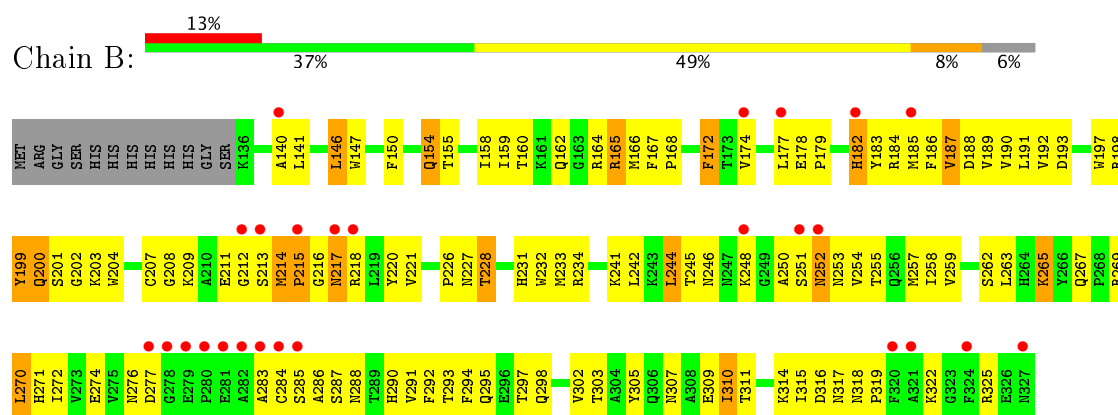
### 3 Residue-property plots

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 ( $\text{RSRZ} > 2$ ). 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.

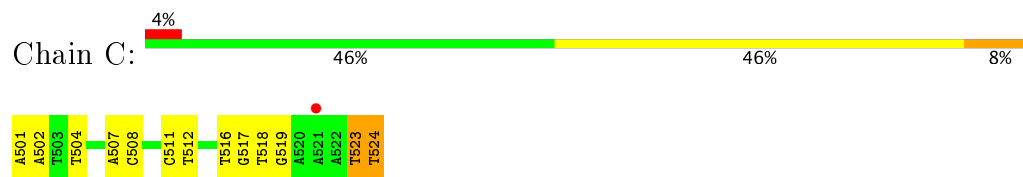
#### • Molecule 1: T-box transcription factor TBX21



#### • Molecule 1: T-box transcription factor TBX21

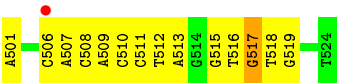


#### • Molecule 2: DNA



#### • Molecule 2: DNA

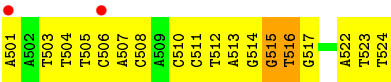




● Molecule 2: DNA



● Molecule 2: DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.45Å 70.45Å 438.39Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.70 – 2.95 43.70 – 2.95	Depositor EDS
% Data completeness (in resolution range)	87.8 (43.70-2.95) 88.0 (43.70-2.95)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.25 (at 2.95Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.266 , 0.294 0.266 , 0.295	Depositor DCC
$R_{free}$ test set	1154 reflections (5.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	81.5	Xtriage
Anisotropy	0.019	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , 68.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.448 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	5062	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	131.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.98% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.63	0/1594	0.95	2/2157 (0.1%)
1	B	0.60	0/1594	0.88	0/2157
2	C	1.05	0/548	1.16	4/844 (0.5%)
2	D	1.03	0/548	1.19	2/844 (0.2%)
2	E	1.22	4/548 (0.7%)	1.18	4/844 (0.5%)
2	F	1.01	1/548 (0.2%)	1.20	2/844 (0.2%)
All	All	0.84	5/5380 (0.1%)	1.04	14/7690 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	523	DT	C1'-N1	8.86	1.60	1.49
2	E	501	DA	N9-C4	6.30	1.41	1.37
2	F	522	DA	C3'-O3'	-5.86	1.36	1.44
2	E	504	DT	C1'-N1	5.41	1.56	1.49
2	E	523	DT	P-O5'	5.16	1.65	1.59

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	214	MET	C-N-CD	-13.72	90.42	120.60
2	E	523	DT	O4'-C1'-N1	8.42	113.89	108.00
1	A	214	MET	C-N-CA	8.00	155.61	122.00
2	D	517	DG	O5'-P-OP2	-6.03	100.28	105.70
2	C	523	DT	OP1-P-O3'	5.60	117.53	105.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	235	GLN	Peptide

## 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	1553	0	1512	114	0
1	B	1553	0	1512	122	0
2	C	489	0	274	11	0
2	D	489	0	274	15	0
2	E	489	0	274	11	0
2	F	489	0	274	17	0
All	All	5062	0	4120	267	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 30.

The worst 5 of 267 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:188:ASP:HB2	1:B:217:ASN:HD21	1.20	1.03
1:A:231:HIS:HA	1:A:234:ARG:HG3	1.41	1.00
1:A:154:GLN:HG3	1:A:204:TRP:CD1	2.04	0.92
1:B:251:SER:C	1:B:252:ASN:HD22	1.75	0.90
1:B:218:ARG:HH21	1:B:259:VAL:HG11	1.39	0.87

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	190/204 (93%)	153 (80%)	30 (16%)	7 (4%)	4	14
1	B	190/204 (93%)	156 (82%)	26 (14%)	8 (4%)	3	11
All	All	380/408 (93%)	309 (81%)	56 (15%)	15 (4%)	3	13

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	214	MET
1	A	215	PRO
1	A	234	ARG
1	B	154	GLN
1	B	213	SER

### 5.3.2 Protein sidechains

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	169/179 (94%)	154 (91%)	15 (9%)	11	32
1	B	169/179 (94%)	153 (90%)	16 (10%)	10	29
All	All	338/358 (94%)	307 (91%)	31 (9%)	11	31

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	310	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	172	PHE
1	B	270	LEU
1	B	146	LEU
1	B	182	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	261	GLN
1	A	276	ASN
1	B	154	GLN
1	A	200	GLN
1	A	290	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	192/204 (94%)	0.65	21 (10%) 6 5	75, 115, 211, 278	0
1	B	192/204 (94%)	0.94	26 (13%) 3 2	76, 115, 211, 304	0
2	C	24/24 (100%)	0.23	1 (4%) 37 34	89, 125, 180, 188	0
2	D	24/24 (100%)	0.23	1 (4%) 37 34	94, 127, 176, 178	0
2	E	24/24 (100%)	0.22	2 (8%) 12 9	100, 134, 178, 200	0
2	F	24/24 (100%)	0.86	2 (8%) 12 9	99, 136, 224, 259	0
All	All	480/504 (95%)	0.71	53 (11%) 6 5	75, 121, 211, 304	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	282	ALA	15.4
1	B	280	PRO	15.4
2	F	501	DA	14.2
1	B	283	ALA	13.2
1	A	282	ALA	8.6

### 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.