



Full wwPDB X-ray Structure Validation Report ⓘ

May 30, 2020 – 01:49 am BST

PDB ID : 6FQP
Title : Crystal structure of TALE homeobox domain transcription factor TGIF1 with its consensus DNA
Authors : Guca, E.; Macias, M.J.
Deposited on : 2018-02-14
Resolution : 2.42 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

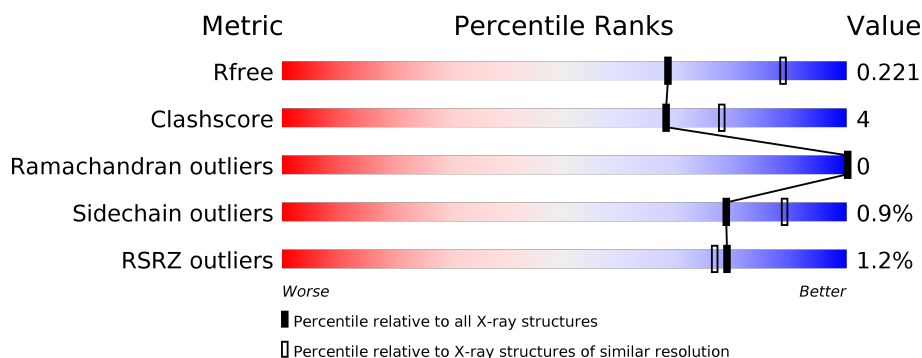
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.42 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}	130704	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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	L	16	<div> <div></div> <div>69%13%19%</div> </div>
1	M	16	<div> <div>6%</div> <div>69%25%6%</div> </div>
2	A	101	<div> <div>%</div> <div>62%...34%</div> </div>
2	B	101	<div> <div></div> <div>57%8%35%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3223 atoms, of which 1415 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 DNA chain called DNA (5'-D(P*AP*TP*TP*GP*AP*CP*AP*GP*CP*TP*GP*TP*CP*AP*AP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	L	16	Total	C	H	N	O	P	2	0	0
			510	157	182	59	96	16			
1	M	16	Total	C	H	N	O	P	2	0	0
			510	157	182	59	96	16			

- Molecule 2 is a protein called Homeobox protein TGIF1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	A	67	Total	C	H	N	O	S	1	0	0
			1063	344	520	102	95	2			
2	B	66	Total	C	H	N	O	S	2	0	0
			1071	346	531	103	89	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	148	GLY	-	expression tag	UNP Q15583
A	149	PRO	-	expression tag	UNP Q15583
B	148	GLY	-	expression tag	UNP Q15583
B	149	PRO	-	expression tag	UNP Q15583

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	L	1	Total	Ca	0	0
			1	1		

- Molecule 4 is water.

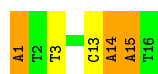
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	L	15	Total 15	O 15	0	0
4	M	19	Total 19	O 19	0	0
4	A	19	Total 19	O 19	0	0
4	B	15	Total 15	O 15	0	0

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.

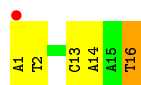
- Molecule 1: DNA (5'-D(P*AP*TP*TP*GP*AP*CP*AP*GP*CP*TP*GP*TP*CP*AP*AP*T)-3')

Chain L: 



- Molecule 1: DNA (5'-D(P*AP*TP*TP*GP*AP*CP*AP*GP*CP*TP*GP*TP*CP*AP*AP*T)-3')

Chain M: 



- Molecule 2: Homeobox protein TGIF1

Chain A: 



- Molecule 2: Homeobox protein TGIF1

Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	50.85Å 67.25Å 102.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.54 – 2.42 56.20 – 2.42	Depositor EDS
% Data completeness (in resolution range)	98.7 (45.54-2.42) 98.9 (56.20-2.42)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.16 (at 2.42Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.205 , 0.241 0.225 , 0.221	Depositor DCC
R_{free} test set	650 reflections (4.72%)	wwPDB-VP
Wilson B-factor (Å ²)	43.5	Xtriage
Anisotropy	0.913	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 41.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	3223	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 20.99 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.7208e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	L	1.44	3/367 (0.8%)	1.30	3/564 (0.5%)
1	M	1.31	2/367 (0.5%)	1.23	2/564 (0.4%)
2	A	0.49	0/555	0.64	0/752
2	B	0.57	0/552	0.70	0/746
All	All	0.96	5/1841 (0.3%)	0.97	5/2626 (0.2%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	L	14	DA	N9-C4	-7.72	1.33	1.37
1	M	1	DA	N9-C4	-6.32	1.34	1.37
1	L	15	DA	N9-C4	-6.29	1.34	1.37
1	L	14	DA	N3-C4	-6.18	1.31	1.34
1	M	16	DT	C1'-N1	5.51	1.56	1.49

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	M	1	DA	O4'-C1'-N9	5.41	111.79	108.00
1	L	15	DA	O5'-P-OP1	-5.25	100.97	105.70
1	L	3	DT	N3-C4-O4	5.11	122.97	119.90
1	M	1	DA	O5'-P-OP1	5.11	116.83	110.70
1	L	1	DA	OP2-P-O3'	5.06	116.33	105.20

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	L	328	182	182	5	0
1	M	328	182	182	3	0
2	A	543	520	519	2	0
2	B	540	531	527	6	0
3	L	1	0	0	0	0
4	A	19	0	0	0	0
4	B	15	0	0	0	0
4	L	15	0	0	0	0
4	M	19	0	0	1	0
All	All	1808	1415	1410	12	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 4.

All (12) 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:L:15:DA:H2	1:M:2:DT:H3	1.50	0.60
1:L:14:DA:O3'	2:B:166:ARG:HG2	2.03	0.58
2:B:201:SER:OG	2:B:206:LEU:O	2.22	0.52
1:L:13:DC:H2''	1:L:14:DA:C8	2.48	0.49
1:L:1:DA:H2	1:M:16:DT:O2	2.00	0.45
2:A:223:LEU:O	2:A:226:MET:HB2	2.17	0.44
1:M:13:DC:H2''	1:M:14:DA:C8	2.52	0.44
2:B:215:PHE:HB3	2:B:219:ARG:NH1	2.34	0.43
1:L:15:DA:OP2	2:B:166:ARG:NH2	2.51	0.43
2:A:228:ARG:C	2:A:230:ASP:H	2.24	0.41
4:M:205:HOH:O	2:B:167:ARG:HD3	2.20	0.41
2:B:200:LEU:O	2:B:204:THR:HG23	2.21	0.40

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	
2	A	65/101 (64%)	62 (95%)	3 (5%)	0	100	100
2	B	64/101 (63%)	57 (89%)	7 (11%)	0	100	100
All	All	129/202 (64%)	119 (92%)	10 (8%)	0	100	100

There are no Ramachandran outliers to report.

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	
2	A	54/90 (60%)	53 (98%)	1 (2%)	57	74
2	B	52/90 (58%)	52 (100%)	0	100	100
All	All	106/180 (59%)	105 (99%)	1 (1%)	78	89

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	230	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

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](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95th 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 ²)	Q<0.9
1	L	16/16 (100%)	-0.20	0 100 100	40, 46, 59, 60	1 (6%)
1	M	16/16 (100%)	-0.07	1 (6%) 20 18	39, 49, 65, 66	1 (6%)
2	A	67/101 (66%)	-0.02	1 (1%) 73 71	34, 46, 79, 93	3 (4%)
2	B	66/101 (65%)	-0.07	0 100 100	36, 44, 66, 84	4 (6%)
All	All	165/234 (70%)	-0.06	2 (1%) 79 76	34, 46, 71, 93	9 (5%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	230	ASP	2.5
1	M	1	DA	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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	CA	L	101	1/1	0.84	0.16	87,87,87,87	0

6.5 Other polymers [i](#)

There are no such residues in this entry.