



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 08:09 am GMT

PDB ID : 4C5I
Title : Crystal structure of MBTD1 YY1 complex
Authors : Alfieri, C.; Glatt, S.; Mueller, C.W.
Deposited on : 2013-09-12
Resolution : 2.59 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

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

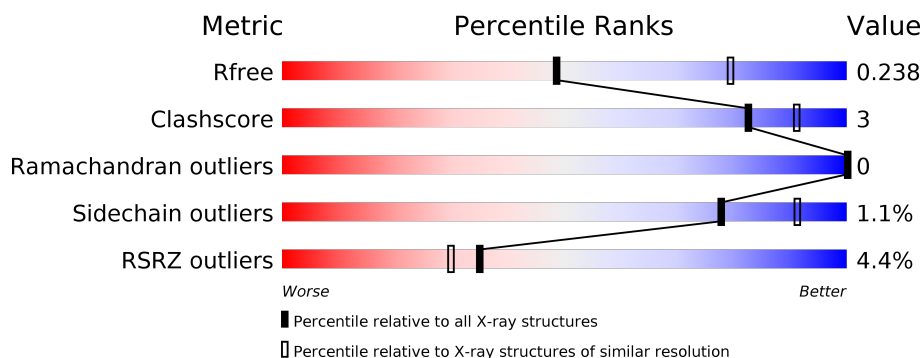
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 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 ≥ 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	440	<div> <div>4%</div> <div> <div></div> <div>86%</div> <div>10%</div> <div>.</div> </div> </div>
1	B	440	<div> <div>3%</div> <div> <div></div> <div>86%</div> <div>8%</div> <div>6%</div> </div> </div>
2	C	30	<div> <div>27%</div> <div> <div></div> <div>40%</div> <div>13%</div> <div>47%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6904 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 MBT DOMAIN-CONTAINING PROTEIN 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	423	Total	C	N	O	S	80	0	0
			3371	2165	576	605	25			
1	B	415	Total	C	N	O	S	72	0	0
			3310	2131	568	587	24			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	127	GLY	-	EXPRESSION TAG	UNP Q05BQ5
A	128	ALA	-	EXPRESSION TAG	UNP Q05BQ5
A	129	MET	-	EXPRESSION TAG	UNP Q05BQ5
B	127	GLY	-	EXPRESSION TAG	UNP Q05BQ5
B	128	ALA	-	EXPRESSION TAG	UNP Q05BQ5
B	129	MET	-	EXPRESSION TAG	UNP Q05BQ5

- Molecule 2 is a protein called TRANSCRIPTIONAL REPRESSOR PROTEIN YY1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	16	Total	C	N	O	S	32	0	0
			135	87	22	25	1			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	56	Total	O	0	0
			56	56		
3	B	32	Total	O	0	0
			32	32		

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	70.36Å 104.71Å 135.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.81 – 2.59 48.81 – 2.59	Depositor EDS
% Data completeness (in resolution range)	99.2 (48.81-2.59) 99.3 (48.81-2.59)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.50 (at 2.58Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.1_1168)	Depositor
R, R_{free}	0.213 , 0.233 0.216 , 0.238	Depositor DCC
R_{free} test set	1588 reflections (5.00%)	DCC
Wilson B-factor (Å ²)	40.2	Xtriage
Anisotropy	0.731	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 44.9	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	0.92	EDS
Total number of atoms	6904	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.20% of the height of the origin peak. No significant pseudotranslation is detected.*

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 [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	0.22	0/3475	0.40	1/4725 (0.0%)
1	B	0.22	0/3414	0.40	1/4643 (0.0%)
2	C	0.22	0/136	0.40	0/180
All	All	0.22	0/7025	0.40	2/9548 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	272	ASP	CB-CG-OD2	5.22	123.00	118.30
1	A	361	ASP	CB-CG-OD2	5.17	122.95	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	3371	0	3257	21	1
1	B	3310	0	3206	17	1
2	C	135	0	135	3	0
3	A	56	0	0	0	0
3	B	32	0	0	0	0
All	All	6904	0	6598	38	1

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 3.

The worst 5 of 38 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:A:324:ASP:HB3	1:A:327:ASP:HB2	1.64	0.79
1:B:324:ASP:HB3	1:B:327:ASP:HB2	1.68	0.75
1:A:408:LYS:HB2	1:A:416:MET:HB2	1.78	0.65
1:B:408:LYS:HB2	1:B:416:MET:HB2	1.80	0.64
1:B:308:GLU:OE1	1:B:316:ARG:NH1	2.33	0.61

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:139:GLU:OE2	1:B:236:ALA:CB[4_446]	2.14	0.06

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	419/440 (95%)	407 (97%)	12 (3%)	0	100	100
1	B	411/440 (93%)	400 (97%)	11 (3%)	0	100	100
2	C	12/30 (40%)	12 (100%)	0	0	100	100
All	All	842/910 (92%)	819 (97%)	23 (3%)	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	364/381 (96%)	359 (99%)	5 (1%)	71	88
1	B	356/381 (93%)	354 (99%)	2 (1%)	89	96
2	C	16/28 (57%)	15 (94%)	1 (6%)	21	40
All	All	736/790 (93%)	728 (99%)	8 (1%)	78	91

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	498	LEU
2	C	214	THR
1	B	269	LEU
1	A	308	GLU
1	A	505	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	214	ASN
2	C	211	GLN

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

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 [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	A	423/440 (96%)	0.32	19 (4%) 34 29	13, 30, 60, 78	28 (6%)
1	B	415/440 (94%)	0.22	11 (2%) 55 50	17, 32, 54, 63	22 (5%)
2	C	16/30 (53%)	2.47	8 (50%) 0 0	58, 64, 72, 73	11 (68%)
All	All	854/910 (93%)	0.32	38 (4%) 35 30	13, 31, 60, 78	61 (7%)

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	219	PHE	5.3
2	C	223	MET	4.4
2	C	224	TRP	3.9
2	C	209	GLN	3.9
1	A	362	GLY	3.6

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.