



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:38 PM GMT

PDB ID : 3UGM
Title : Structure of TAL effector PthXo1 bound to its DNA target
Authors : Mak, A.N.S.; Bradley, P.; Cernadas, R.A.; Bogdanove, A.J.; Stoddard, B.L.
Deposited on : 2011-11-02
Resolution : 3.00 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
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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) : trunk26865

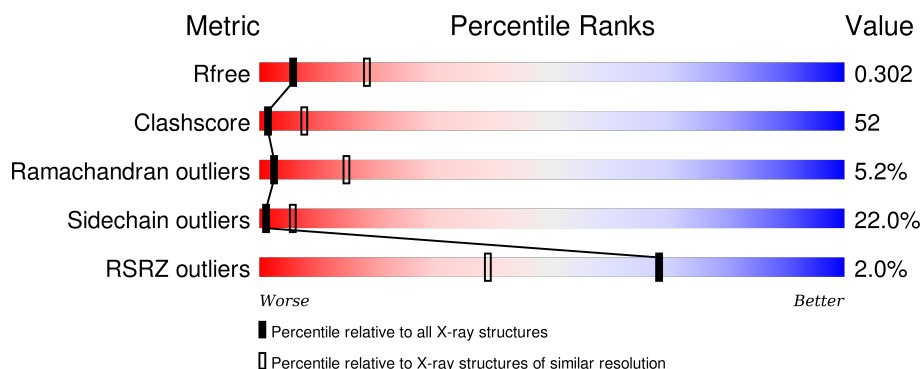
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	1047	<div> <div>25%</div> <div>41%</div> <div>14%</div> <div>•</div> <div>18%</div> </div>
2	B	38	<div> <div>8%</div> <div>26%</div> <div>61%</div> <div>13%</div> </div>
3	C	38	<div> <div>3%</div> <div>32%</div> <div>58%</div> <div>11%</div> </div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 7853 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 TAL effector AvrBs3/PthA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	854	Total	C	N	O	S	0	0	0
			6085	3792	1123	1147	23			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	MET	-	EXPRESSION TAG	UNP B2SU53
A	104	ALA	-	EXPRESSION TAG	UNP B2SU53
A	105	SER	-	EXPRESSION TAG	UNP B2SU53
A	106	SER	-	EXPRESSION TAG	UNP B2SU53
A	107	HIS	-	EXPRESSION TAG	UNP B2SU53
A	108	HIS	-	EXPRESSION TAG	UNP B2SU53
A	109	HIS	-	EXPRESSION TAG	UNP B2SU53
A	110	HIS	-	EXPRESSION TAG	UNP B2SU53
A	111	HIS	-	EXPRESSION TAG	UNP B2SU53
A	112	HIS	-	EXPRESSION TAG	UNP B2SU53
A	113	SER	-	EXPRESSION TAG	UNP B2SU53
A	114	SER	-	EXPRESSION TAG	UNP B2SU53
A	115	GLY	-	EXPRESSION TAG	UNP B2SU53
A	116	LEU	-	EXPRESSION TAG	UNP B2SU53
A	117	VAL	-	EXPRESSION TAG	UNP B2SU53
A	118	PRO	-	EXPRESSION TAG	UNP B2SU53
A	119	ARG	-	EXPRESSION TAG	UNP B2SU53
A	120	GLY	-	EXPRESSION TAG	UNP B2SU53
A	121	SER	-	EXPRESSION TAG	UNP B2SU53
A	122	SER	-	EXPRESSION TAG	UNP B2SU53
A	123	GLY	-	EXPRESSION TAG	UNP B2SU53
A	124	SER	-	EXPRESSION TAG	UNP B2SU53
A	125	SER	-	EXPRESSION TAG	UNP B2SU53
A	126	MET	-	EXPRESSION TAG	UNP B2SU53
A	173	ARG	GLY	ENGINEERED MUTATION	UNP B2SU53
A	198	GLN	ARG	ENGINEERED MUTATION	UNP B2SU53
A	209	THR	LYS	ENGINEERED MUTATION	UNP B2SU53

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	212	HIS	ASP	ENGINEERED MUTATION	UNP B2SU53
A	213	ILE	MET	ENGINEERED MUTATION	UNP B2SU53
A	215	THR	ALA	ENGINEERED MUTATION	UNP B2SU53
A	244	ASP	VAL	ENGINEERED MUTATION	UNP B2SU53
A	272	MET	VAL	ENGINEERED MUTATION	UNP B2SU53

- Molecule 2 is a DNA chain called DNA-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	38	Total	C	N	O	P	0	0	0
			764	367	137	223	37			

- Molecule 3 is a DNA chain called DNA-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	38	Total	C	N	O	P	0	0	0
			788	376	146	229	37			

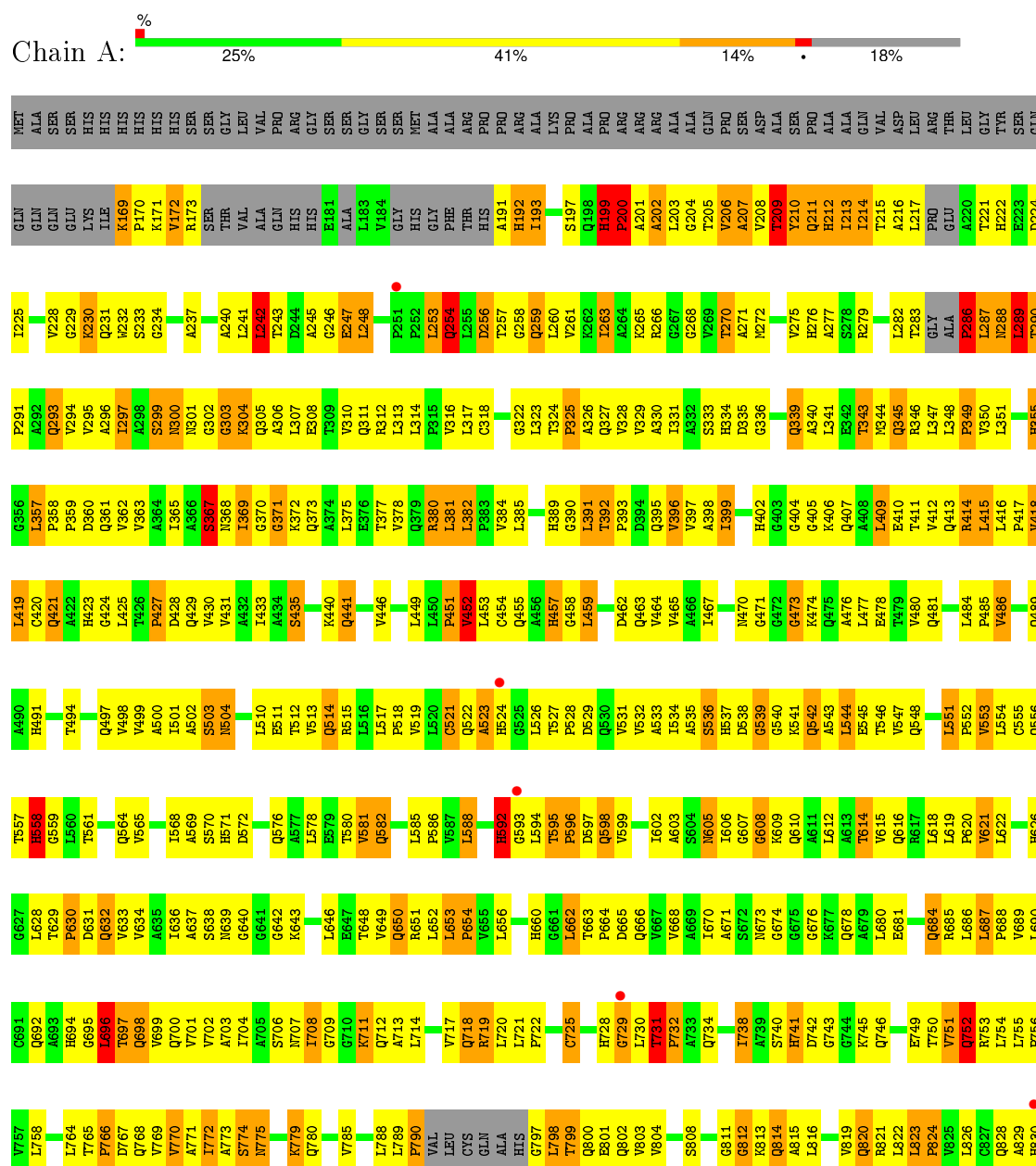
- Molecule 4 is water.

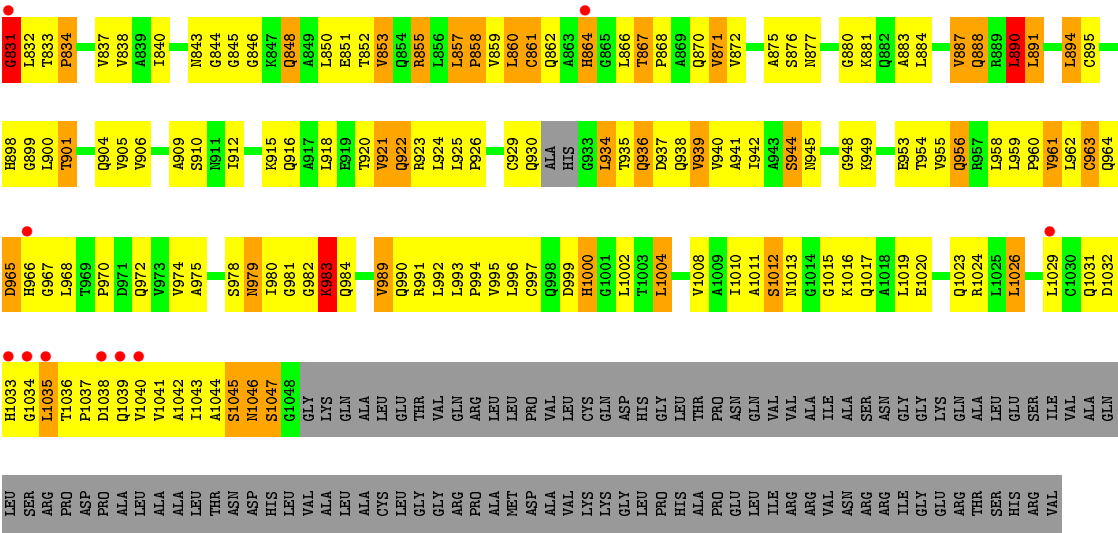
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	159	Total	O	0	0
			159	159		
4	B	23	Total	O	0	0
			23	23		
4	C	34	Total	O	0	0
			34	34		

3 Residue-property plots

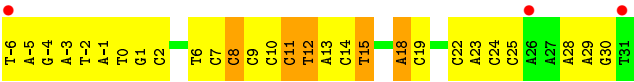
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: TAL effector AvrBs3/PthA





• Molecule 2: DNA-1



• Molecule 3: DNA-2



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	95.58 Å 248.48 Å 54.65 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00 44.60 – 3.01	Depositor EDS
% Data completeness (in resolution range)	95.3 (50.00-3.00) 95.9 (44.60-3.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.01 (at 3.01 Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.266 , 0.294 0.269 , 0.302	Depositor DCC
R_{free} test set	1326 reflections (5.15%)	DCC
Wilson B-factor (Å ²)	76.4	Xtriage
Anisotropy	0.259	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 105.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 25725 reflections	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	7853	wwPDB-VP
Average B, all atoms (Å ²)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.50% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	1.07	42/6153 (0.7%)	1.18	64/8397 (0.8%)
2	B	1.44	4/855 (0.5%)	1.22	2/1314 (0.2%)
3	C	1.36	6/885 (0.7%)	1.11	0/1368
All	All	1.15	52/7893 (0.7%)	1.18	66/11079 (0.6%)

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	9

The worst 5 of 52 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	191	ALA	CA-CB	-15.15	1.20	1.52
1	A	169	LYS	CA-CB	-14.67	1.21	1.53
1	A	210	TYR	CE2-CZ	-10.88	1.24	1.38
1	A	212	HIS	CA-CB	-10.45	1.30	1.53
1	A	210	TYR	CG-CD1	-9.62	1.26	1.39

The worst 5 of 66 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	201	ALA	CB-CA-C	9.49	124.34	110.10
1	A	202	ALA	N-CA-CB	-8.41	98.33	110.10
1	A	774	SER	CB-CA-C	-7.60	95.66	110.10
1	A	288	ASN	C-N-CA	7.46	140.35	121.70
1	A	559	GLY	N-CA-C	7.44	131.70	113.10

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	169	LYS	Peptide
1	A	200	PRO	Mainchain
1	A	209	THR	Peptide
1	A	211	GLN	Peptide
1	A	215	THR	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	6085	0	6242	687	0
2	B	764	0	427	52	0
3	C	788	0	432	49	0
4	A	159	0	0	88	0
4	B	23	0	0	12	0
4	C	34	0	0	12	0
All	All	7853	0	7101	772	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 52.

The worst 5 of 772 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:454:CYS:HA	1:A:457:HIS:O	1.26	1.27
1:A:729:GLY:O	1:A:730:LEU:HD23	1.14	1.26
1:A:1008:VAL:HG11	4:A:101:HOH:O	1.37	1.21
1:A:684:GLN:HB2	4:A:1175:HOH:O	1.43	1.16
1:A:389:HIS:CE1	1:A:416:LEU:HD23	1.79	1.16

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	839/1047 (80%)	549 (65%)	246 (29%)	44 (5%)	2 15

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	200	PRO
1	A	303	GLY
1	A	371	GLY
1	A	1034	GLY
1	A	221	THR

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	635/816 (78%)	495 (78%)	140 (22%)	1 6

5 of 140 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	561	THR
1	A	684	GLN
1	A	980	ILE
1	A	580	THR
1	A	595	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	537	HIS
1	A	592	HIS
1	A	945	ASN
1	A	582	GLN
1	A	610	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 [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, 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	854/1047 (81%)	-0.21	15 (1%) 71 43	27, 84, 153, 200	0
2	B	38/38 (100%)	0.11	3 (7%) 15 5	31, 46, 132, 137	0
3	C	38/38 (100%)	0.05	1 (2%) 59 29	43, 70, 129, 135	0
All	All	930/1123 (82%)	-0.18	19 (2%) 68 39	27, 83, 153, 200	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1040	VAL	4.8
1	A	593	GLY	4.5
2	B	31	DT	4.1
1	A	966	HIS	4.1
1	A	1033	HIS	3.8

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.