W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 17, 2024 – 07:06 PM EST

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)

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

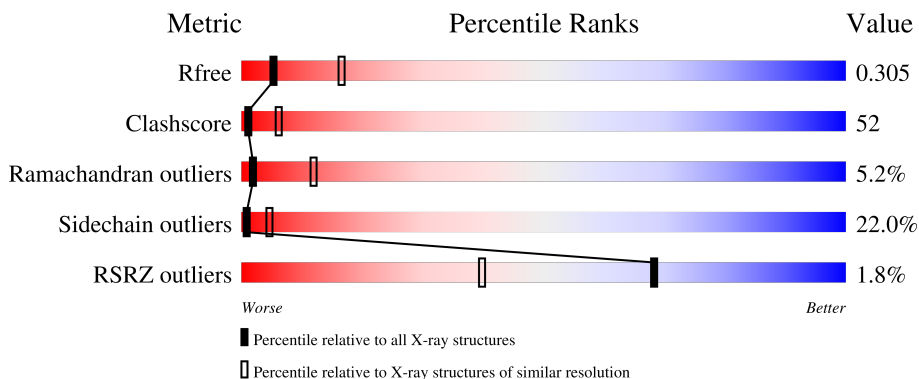|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
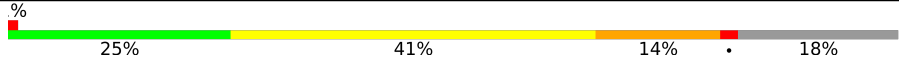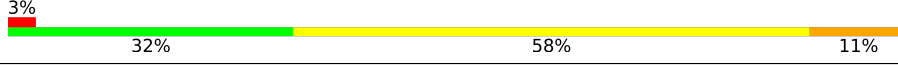*X-RAY DIFFRACTION*

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	Percentile Ranks	Value
Rfree		0.305
Clashscore		52
Ramachandran outliers		5.2%
Sidechain outliers		22.0%
RSRZ outliers		1.8%

Worse   *Better*

▌ Percentile relative to all X-ray structures

⫾ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (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 of 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 <=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	% 25% 41% 14% • 18%
2	B	38	5% 26% 61% 13%
3	C	38	3% 32% 58% 11%

2 Entry composition ⓘ

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 6085	C 3792	N 1123	O 1147	S 23	0	0	0

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 764	C 367	N 137	O 223	P 37	0	0	0

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

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

- Molecule 4 is water.

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

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: TAL effector AvrBs3/PthA

G831 L832 T833 P834 V837 V838 A839 I840 N843 G844 G845 G846 K847 Q848 A849 L850 E851 T852 V853 Q854 R855 L856 L857 P858 V859 L860 C861 Q862 A863 H864 G865 L866 T867 P868 A869 Q870 V871 V872 A875 S876 N877 G880 K881 Q882 A883 L884 V887 Q888 R889 L890 L891 L894 C895

H898 G899 L900 T901 Q904 V905 V906 A909 S910 N911 I912 K915 Q916 A917 L918 E919 T920 V921 Q922 R923 L924 L925 P926 C929 ALA HIS G933 L934 T935 Q936 D937 Q938 V939 V940 A941 I942 A943 S944 N945 G948 K949 E953 T954 V955 Q956 R957 L958 L959 P960 V961 L962 C963 Q964

D965 H966 G967 L968 T969 P970 D971 Q972 V973 V974 A975 S978 N979 I980 G981 G982 K983 Q984 V989 Q990 R991 L992 L993 P994 V995 L996 C997 Q998 D999 H1000 G1001 L1002 T1003 L1004 V1008 A1009 I1010 A1011 S1012 N1013 G1014 G1015 K1016 Q1017 A1018 L1019 E1020 Q1023 R1024 L1025 L1026 L1029 C1030 Q1031 D1032

H1033 G1034 L1035 T1036 P1037 D1038 Q1039 V1040 V1041 A1042 I1043 A1044 S1045 N1046 S1047 G1048 GLY LYS GLN ALA LEU GLU THR VAL GLN ARG LEU LEU PRO VAL LEU CYS GLN ASP HIS GLY LEU THR PRO ASN GLN VAL VAL ALA ILE ALA SER ASN GLY GLY LYS GLN ALA LEU GLU SER ILE VAL ALA GLN

LEU SER ARG PRO ASP PRO ALA LEU ALA ALA LEU THR ASN ASP HIS LEU VAL ALA LEU ALA CYS LEU GLY GLY ARG PRO ALA MET ASP ALA LYS LYS LYS GLY LEU PRO HIS ALA PRO GLU LEU ILE ARG ARG VAL ASN ARG ARG ILE GLU GLU ILE ARG THR SER HIS ARG VAL

• **Molecule 2: DNA-1**

5%
Chain B:
26%   61%   13%

T-6 A-5 G-4 A-3 T-2 A-1 T0 G1 C2 T6 C7 C8 C9 C10 C11 T12 A13 C14 T15 A18 C19 C22 A23 C24 C25 A28 A29 G30 T31

• **Molecule 3: DNA-2**

3%
Chain C:
32%   58%   11%

T1 T2 T3 T4 G11 T12 A13 C14 A15 G16 G19 G20 G21 G22 G23 A24 G25 A26 T27 G28 C29 A30 T31 A32 T33 C34 T35 A36 A37 C38

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	95.58Å 248.48Å 54.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00	Depositor
	44.60 – 3.01	EDS
% Data completeness	95.3 (50.00-3.00)	Depositor
(in resolution range)	95.9 (44.60-3.01)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.01 (at 3.01Å)	Xtriage
Refinement program	REFMAC, CNS	Depositor
R, R_{free}	0.266 , 0.294	Depositor
	0.267 , 0.305	DCC
R_{free} test set	1326 reflections (5.15%)	wwPDB-VP
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
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	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 $< |L| >$, $< L^2 >$ 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	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 ⓘ

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	839/1047 (80%)	549 (65%)	246 (29%)	44 (5%)	2	12

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 ⓘ

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	4

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

Mol	Chain	Res	Type
1	A	891	LEU
1	A	922	GLN
1	A	983	LYS
1	A	452	VAL
1	A	441	GLN

Sometimes 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	904	GLN
1	A	945	ASN
1	A	504	ASN
1	A	475	GLN
1	A	1023	GLN

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 monosaccharides 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, 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%)	-0.23	14 (1%)	72	44	27, 84, 153, 200	0
2	B	38/38 (100%)	0.08	2 (5%)	26	10	31, 46, 132, 137	0
3	C	38/38 (100%)	0.02	1 (2%)	56	27	43, 70, 129, 135	0
All	All	930/1123 (82%)	-0.20	17 (1%)	68	40	27, 83, 153, 200	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1040	VAL	4.7
1	A	593	GLY	4.4
1	A	966	HIS	4.0
2	B	31	DT	3.9
1	A	1033	HIS	3.7

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.