



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 6, 2020 – 08:02 PM BST

PDB ID : 6KD6
Title : Shuguo PWWP domain
Authors : Liu, Y.C.; Huang, Y.
Deposited on : 2019-07-01
Resolution : 1.58 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

<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	:	2.13.1
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)	:	2.13.1

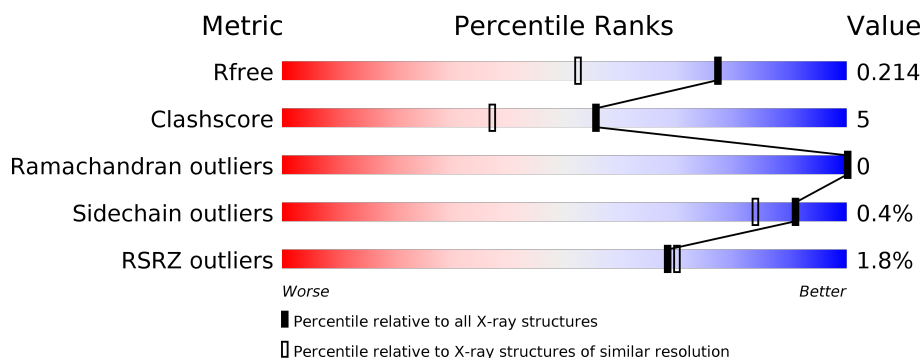
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.58 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	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

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 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	A	85	<div> <div>5%</div> <div> <div></div> <div>87%</div> <div>7%</div> <div>6%</div> </div> </div>
1	B	85	<div> <div>%</div> <div> <div></div> <div>84%</div> <div>15%</div> <div>.</div> </div> </div>
1	C	85	<div> <div>%</div> <div> <div></div> <div>84%</div> <div>12%</div> <div>5%</div> </div> </div>
1	D	85	<div> <div></div> <div> <div></div> <div>86%</div> <div>8%</div> <div>6%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3128 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 LD23804p.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	80	Total	C	N	O	S	0	0	0
			645	423	103	118	1			
1	B	84	Total	C	N	O	S	0	0	0
			674	438	111	124	1			
1	C	81	Total	C	N	O	S	0	0	0
			656	429	107	119	1			
1	D	80	Total	C	N	O	S	0	0	0
			647	424	105	117	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	7	SER	-	expression tag	UNP Q9VAA9
B	7	SER	-	expression tag	UNP Q9VAA9
C	7	SER	-	expression tag	UNP Q9VAA9
D	7	SER	-	expression tag	UNP Q9VAA9


- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	102	Total	O	0	0
			102	102		
2	B	140	Total	O	0	0
			140	140		
2	C	126	Total	O	0	0
			126	126		
2	D	138	Total	O	0	0
			138	138		


i

- Molecule 1: LD23804p



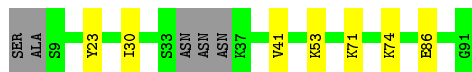
- Chain B:  84% 15% 1%



- Chain C:  84% 12% 5%



- Chain D: 86% 8% 6%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	49.87Å 51.13Å 58.73Å 90.00° 93.50° 90.00°	Depositor
Resolution (Å)	39.14 – 1.58 39.14 – 1.58	Depositor EDS
% Data completeness (in resolution range)	99.3 (39.14-1.58) 99.1 (39.14-1.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.03 (at 1.58Å)	Xtriage
Refinement program	PHENIX 1.13 _2998	Depositor
R, R_{free}	0.177 , 0.214 0.177 , 0.214	Depositor DCC
R_{free} test set	2025 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	16.2	Xtriage
Anisotropy	0.470	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 39.2	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	0.96	EDS
Total number of atoms	3128	wwPDB-VP
Average B, all atoms (Å ²)	21.0	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 53.18 % 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 4.3736e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² 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	0.32	0/660	0.48	0/887
1	B	0.36	0/689	0.49	0/925
1	C	0.36	0/671	0.51	0/901
1	D	0.38	0/662	0.51	0/888
All	All	0.35	0/2682	0.50	0/3601

There are no bond length outliers.

There are no bond angle outliers.

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	645	0	651	4	0
1	B	674	0	678	12	0
1	C	656	0	664	10	0
1	D	647	0	656	6	0
2	A	102	0	0	1	2
2	B	140	0	0	7	0
2	C	126	0	0	5	0
2	D	138	0	0	2	0
All	All	3128	0	2649	27	2

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 5.

The worst 5 of 27 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:37:LYS:N	2:A:101:HOH:O	1.99	0.93
1:C:74:LYS:NZ	2:C:101:HOH:O	2.05	0.88
1:D:86:GLU:OE2	2:D:101:HOH:O	2.11	0.69
1:C:64:LYS:HE3	2:C:145:HOH:O	2.01	0.60
1:A:78:PHE:CE1	1:B:71:LYS:HE3	2.38	0.58

All (2) 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)
2:A:104:HOH:O	2:A:154:HOH:O[2_546]	1.85	0.35
2:A:104:HOH:O	2:A:129:HOH:O[2_546]	2.08	0.12

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	76/85 (89%)	76 (100%)	0	0	100	100
1	B	80/85 (94%)	80 (100%)	0	0	100	100
1	C	77/85 (91%)	77 (100%)	0	0	100	100
1	D	76/85 (89%)	76 (100%)	0	0	100	100
All	All	309/340 (91%)	309 (100%)	0	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	67/71 (94%)	67 (100%)	0	100	100
1	B	70/71 (99%)	70 (100%)	0	100	100
1	C	68/71 (96%)	67 (98%)	1 (2%)	65	42
1	D	67/71 (94%)	67 (100%)	0	100	100
All	All	272/284 (96%)	271 (100%)	1 (0%)	91	84

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

Mol	Chain	Res	Type
1	C	66	ARG

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

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 monosaccharides 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	80/85 (94%)	0.16	4 (5%) 28 29	11, 22, 35, 42	0
1	B	84/85 (98%)	-0.12	1 (1%) 79 80	11, 17, 27, 35	0
1	C	81/85 (95%)	0.02	1 (1%) 79 80	12, 20, 33, 40	0
1	D	80/85 (94%)	-0.20	0 100 100	11, 16, 24, 31	0
All	All	325/340 (95%)	-0.03	6 (1%) 68 70	11, 19, 32, 42	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	89	LEU	3.0
1	B	36	ASN	2.6
1	A	37	LYS	2.6
1	C	89	LEU	2.5
1	A	69	THR	2.4

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 monosaccharides 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.