Full wwPDB X-ray Structure Validation Report ⓘ

Jun 17, 2024 – 09:29 PM EDT

PDB ID	:	5YUG
Title	:	AtVAL1 PHD-Like domain in the P31 space group
Authors	:	Zhang, M.M.; Wu, B.X.
Deposited on	:	2017-11-22
Resolution	:	1.57 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

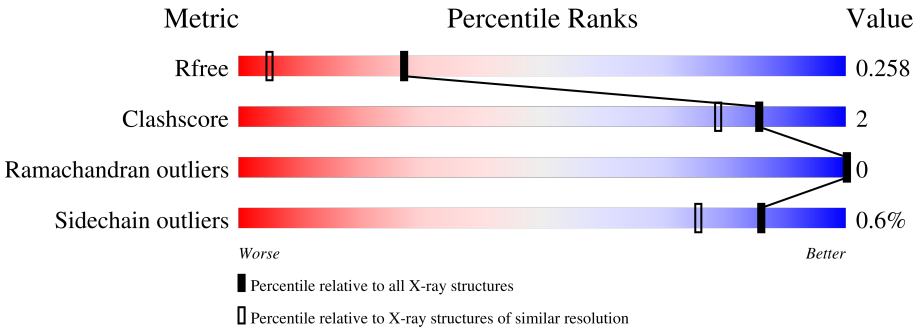
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.57 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)

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%

Mol	Chain	Length	Quality of chain
1	A	110	80% · 16%
1	B	110	78% 5% 16%
1	E	110	90% · 7%
1	G	110	92% · 7%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3127 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.

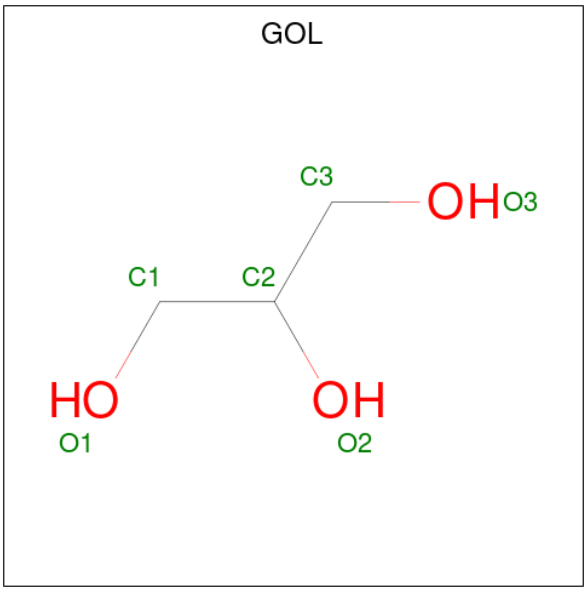
- Molecule 1 is a protein called B3 domain-containing transcription repressor VAL1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	92	Total	C	N	O	S	0	1	0
			705	434	121	134	16			
1	B	92	Total	C	N	O	S	0	0	0
			697	430	120	131	16			
1	E	102	Total	C	N	O	S	0	0	0
			781	479	139	147	16			
1	G	102	Total	C	N	O	S	0	0	0
			781	479	139	147	16			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Zn	0	0
			3	3		
2	B	3	Total	Zn	0	0
			3	3		
2	E	3	Total	Zn	0	0
			3	3		
2	G	3	Total	Zn	0	0
			3	3		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	31	Total O 31 31	0	0
4	B	39	Total O 39 39	0	0
4	E	36	Total O 36 36	0	0
4	G	33	Total O 33 33	0	0

3 Residue-property plots (i)

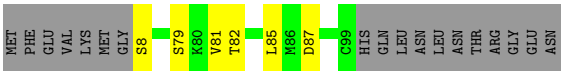
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded 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. 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: B3 domain-containing transcription repressor VAL1

Chain A:

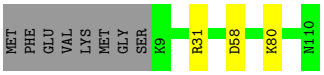
• Molecule 1: B3 domain-containing transcription repressor VAL1

Chain B:



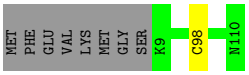
• Molecule 1: B3 domain-containing transcription repressor VAL1

Chain E:



• Molecule 1: B3 domain-containing transcription repressor VAL1

Chain G:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	52.23Å 52.23Å 121.67Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	26.11 – 1.57 26.11 – 1.57	Depositor EDS
% Data completeness (in resolution range)	99.6 (26.11-1.57) 99.6 (26.11-1.57)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.72 (at 1.57Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.220 , 0.254 0.222 , 0.258	Depositor DCC
R_{free} test set	2595 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	12.0	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 25.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.40$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.477 for -h,-k,l 0.336 for h,-h-k,-l 0.336 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3127	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% 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

5.1 Standard geometry

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

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	0.38	0/719	0.58	0/965
1	B	0.36	0/711	0.58	0/954
1	E	0.37	0/796	0.56	0/1069
1	G	0.38	0/796	0.56	0/1069
All	All	0.37	0/3022	0.57	0/4057

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

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	705	0	662	3	0
1	B	697	0	659	10	0
1	E	781	0	738	7	0
1	G	781	0	738	0	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
2	E	3	0	0	0	0
2	G	3	0	0	0	0
3	A	6	0	8	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	6	0	8	1	0
4	A	31	0	0	1	0
4	B	39	0	0	1	0
4	E	36	0	0	0	0
4	G	33	0	0	0	0
All	All	3127	0	2813	13	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 2.

All (13) 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:B:85:LEU:HD22	1:E:80:LYS:HE2	1.85	0.58
1:B:87:ASP:CB	1:E:31:ARG:HH21	2.23	0.52
1:A:88:TYR:N	4:A:301:HOH:O	2.40	0.52
1:B:87:ASP:HB2	1:E:31:ARG:NH2	2.26	0.50
1:B:87:ASP:OD1	1:E:31:ARG:HD3	2.11	0.50
1:B:87:ASP:CG	1:E:31:ARG:HH21	2.19	0.46
1:B:8:SER:N	4:B:304:HOH:O	2.48	0.46
1:B:87:ASP:HB2	1:E:31:ARG:HH21	1.80	0.46
1:A:81:VAL:HG13	1:A:82:THR:HG23	1.99	0.45
1:B:81:VAL:HG13	1:B:82:THR:HG23	1.99	0.43
1:B:79:SER:OG	3:B:204:GOL:H32	2.19	0.42
1:A:79:SER:HA	3:A:204:GOL:H31	2.02	0.41
1:B:87:ASP:CG	1:E:31:ARG:HD3	2.42	0.41

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	91/110 (83%)	90 (99%)	1 (1%)	0	100	100
1	B	90/110 (82%)	89 (99%)	1 (1%)	0	100	100
1	E	100/110 (91%)	98 (98%)	2 (2%)	0	100	100
1	G	100/110 (91%)	98 (98%)	2 (2%)	0	100	100
All	All	381/440 (87%)	375 (98%)	6 (2%)	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	79/94 (84%)	79 (100%)	0	100	100
1	B	78/94 (83%)	78 (100%)	0	100	100
1	E	87/94 (93%)	86 (99%)	1 (1%)	73	55
1	G	87/94 (93%)	86 (99%)	1 (1%)	73	55
All	All	331/376 (88%)	329 (99%)	2 (1%)	86	76

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	58	ASP
1	G	98	CYS

Sometimes 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)

Of 14 ligands modelled in this entry, 12 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	GOL	B	204	-	5,5,5	0.31	0	5,5,5	1.14	0
3	GOL	A	204	-	5,5,5	0.28	0	5,5,5	1.10	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	B	204	-	-	3/4/4/4	-
3	GOL	A	204	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (7) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	204	GOL	O1-C1-C2-C3
3	B	204	GOL	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	B	204	GOL	O1-C1-C2-C3
3	A	204	GOL	O1-C1-C2-O2
3	A	204	GOL	O2-C2-C3-O3
3	B	204	GOL	O2-C2-C3-O3
3	A	204	GOL	C1-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	204	GOL	1	0
3	A	204	GOL	1	0

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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.