Full wwPDB X-ray Structure Validation Report (i)

Aug 30, 2023 – 12:34 AM EDT

PDB ID	:	3P4F
Title	:	Structural and biochemical insights into MLL1 core complex assembly and regulation.
Authors	:	Avdic, V.; Zhang, P.; Lanouette, S.; Groulx, A.; Tremblay, V.; Brunzelle, J.B.; Couture, J.-F.
Deposited on	:	2010-10-06
Resolution	:	2.35 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 (i) 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 (1)) were used in the production of this report:

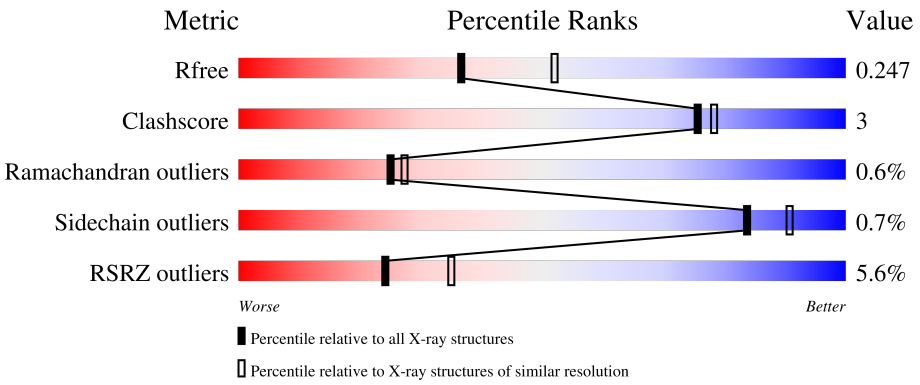
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	317	5%, 88%, 8%, .
2	B	11	9%, 82%, 18%
3	C	10	80%, 10%, 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2718 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 WD repeat-containing protein 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	304	2345	1496	389	450	10	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	18	GLY	-	expression tag	UNP P61964
A	19	ALA	-	expression tag	UNP P61964
A	20	MET	-	expression tag	UNP P61964
A	21	GLY	-	expression tag	UNP P61964

- Molecule 2 is a protein called Retinoblastoma-binding protein 5.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	9	73	45	9	19	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	381	TYR	ASP	engineered mutation	UNP Q15291

- Molecule 3 is a protein called Histone-lysine N-methyltransferase MLL.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	9	56	33	12	11	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	3763	ALA	SER	engineered mutation	UNP Q03164

- Molecule 4 is water.

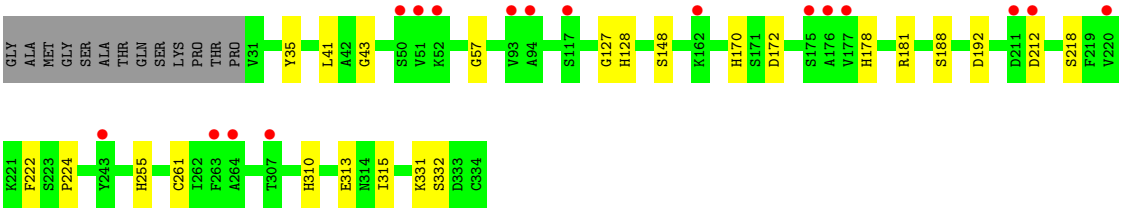
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	221	221	221	0	0
			Total	O		
4	B	12	12	12	0	0
			Total	O		
4	C	11	11	11	0	0

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 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: WD repeat-containing protein 5

Chain A:

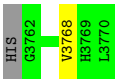


• Molecule 2: Retinoblastoma-binding protein 5

Chain B:

• Molecule 3: Histone-lysine N-methyltransferase MLL

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	53.85Å 81.35Å 81.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.57 – 2.35 32.57 – 2.35	Depositor EDS
% Data completeness (in resolution range)	(Not available) (32.57-2.35) 99.2 (32.57-2.35)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.34 (at 2.34Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, R_{free}	0.194 , 0.249 0.187 , 0.247	Depositor DCC
R_{free} test set	771 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	41.7	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 49.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	0.005 for -h,l,k	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2718	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.17% 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	0.55	0/2401	0.78	0/3260
2	B	0.50	0/73	0.93	0/98
3	C	0.48	0/55	0.80	0/73
All	All	0.55	0/2529	0.78	0/3431

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	2345	0	2302	15	0
2	B	73	0	63	0	0
3	C	56	0	49	0	0
4	A	221	0	0	0	0
4	B	12	0	0	0	0
4	C	11	0	0	0	0
All	All	2718	0	2414	15	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 3.

All (15) 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:310:HIS:HD2	1:A:313:GLU:H	1.28	0.78
1:A:310:HIS:CD2	1:A:313:GLU:H	2.11	0.66
1:A:178:HIS:CD2	1:A:222:PHE:H	2.14	0.65
1:A:315:ILE:HG12	1:A:331:LYS:HG2	1.83	0.60
1:A:178:HIS:HD2	1:A:222:PHE:H	1.49	0.58
1:A:170:HIS:HE1	1:A:188:SER:OG	1.93	0.51
1:A:172:ASP:HB2	1:A:192:ASP:HB3	1.95	0.47
1:A:255:HIS:H	1:A:255:HIS:CD2	2.32	0.47
1:A:181:ARG:HG3	1:A:224:PRO:O	2.15	0.46
1:A:41:LEU:HD12	1:A:41:LEU:N	2.31	0.46
1:A:128:HIS:CE1	1:A:148:SER:HB3	2.51	0.46
1:A:41:LEU:N	1:A:41:LEU:CD1	2.80	0.45
1:A:218:SER:HB2	1:A:261:CYS:HA	1.98	0.45
1:A:57:GLY:O	1:A:310:HIS:HE1	2.02	0.41
1:A:35:TYR:CG	1:A:332:SER:HB2	2.56	0.40

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	302/317 (95%)	288 (95%)	12 (4%)	2 (1%)	22 23
2	B	7/11 (64%)	7 (100%)	0	0	100 100
3	C	7/10 (70%)	6 (86%)	1 (14%)	0	100 100
All	All	316/338 (94%)	301 (95%)	13 (4%)	2 (1%)	25 27

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	43	GLY
1	A	127	GLY

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	261/275 (95%)	260 (100%)	1 (0%)	91 95
2	B	9/11 (82%)	9 (100%)	0	100 100
3	C	3/6 (50%)	2 (67%)	1 (33%)	0 0
All	All	273/292 (94%)	271 (99%)	2 (1%)	84 91

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

Mol	Chain	Res	Type
1	A	212	ASP
3	C	3768	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	170	HIS
1	A	178	HIS
1	A	255	HIS
1	A	265	ASN
1	A	310	HIS
1	A	323	ASN

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	304/317 (95%)	-0.03	17 (5%) 24 35	29, 39, 53, 80	2 (0%)
2	B	9/11 (81%)	0.27	1 (11%) 5 8	41, 44, 70, 87	0
3	C	9/10 (90%)	0.20	0 100 100	35, 41, 51, 67	0
All	All	322/338 (95%)	-0.01	18 (5%) 24 35	29, 40, 54, 87	2 (0%)

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	51	VAL	4.8
1	A	243	TYR	4.2
1	A	177	VAL	3.5
1	A	211	ASP	3.1
1	A	50	SER	3.0
1	A	176	ALA	2.7
1	A	220	VAL	2.6
1	A	263	PHE	2.6
1	A	307	THR	2.6
1	A	162	LYS	2.3
1	A	52	LYS	2.3
1	A	175	SER	2.3
1	A	94	ALA	2.2
1	A	212	ASP	2.1
1	A	93	VAL	2.1
1	A	264	ALA	2.1
1	A	117	SER	2.1
2	B	373	GLU	2.0

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.