



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:43 am BST

PDB ID : 5VK3
Title : Apo ctPRC2 with E840A and K852D mutations in Ezh2
Authors : Bratkowski, M.A.; Liu, X.
Deposited on : 2017-04-20
Resolution : 2.11 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	:	2.11
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)	:	2.11

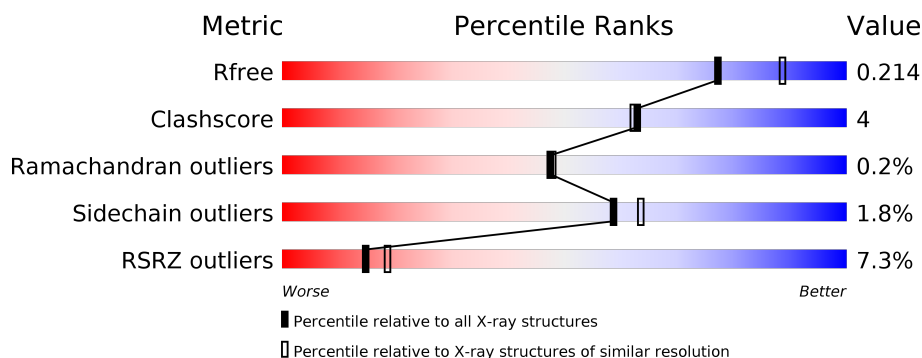
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.11 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	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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	605	<div> <div>5%</div> <div> <div></div> <div>70%</div> <div>7%</div> <div>23%</div> </div> </div>
2	B	937	<div> <div>7%</div> <div> <div></div> <div>75%</div> <div>9%</div> <div>15%</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10990 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 Polycomb Protein EED.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	468	Total	C	N	O	S	0	2	0
			3692	2359	637	677	19			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-39	MET	-	initiating methionine	UNP G0S8H7
A	-38	ALA	-	expression tag	UNP G0S8H7
A	-37	SER	-	expression tag	UNP G0S8H7
A	-36	ALA	-	expression tag	UNP G0S8H7
A	-35	TRP	-	expression tag	UNP G0S8H7
A	-34	SER	-	expression tag	UNP G0S8H7
A	-33	HIS	-	expression tag	UNP G0S8H7
A	-32	PRO	-	expression tag	UNP G0S8H7
A	-31	GLN	-	expression tag	UNP G0S8H7
A	-30	PHE	-	expression tag	UNP G0S8H7
A	-29	GLU	-	expression tag	UNP G0S8H7
A	-28	LYS	-	expression tag	UNP G0S8H7
A	-27	GLY	-	expression tag	UNP G0S8H7
A	-26	GLY	-	expression tag	UNP G0S8H7
A	-25	GLY	-	expression tag	UNP G0S8H7
A	-24	SER	-	expression tag	UNP G0S8H7
A	-23	GLY	-	expression tag	UNP G0S8H7
A	-22	GLY	-	expression tag	UNP G0S8H7
A	-21	GLY	-	expression tag	UNP G0S8H7
A	-20	SER	-	expression tag	UNP G0S8H7
A	-19	GLY	-	expression tag	UNP G0S8H7
A	-18	GLY	-	expression tag	UNP G0S8H7
A	-17	SER	-	expression tag	UNP G0S8H7
A	-16	ALA	-	expression tag	UNP G0S8H7
A	-15	TRP	-	expression tag	UNP G0S8H7
A	-14	SER	-	expression tag	UNP G0S8H7
A	-13	HIS	-	expression tag	UNP G0S8H7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	PRO	-	expression tag	UNP G0S8H7
A	-11	GLN	-	expression tag	UNP G0S8H7
A	-10	PHE	-	expression tag	UNP G0S8H7
A	-9	GLU	-	expression tag	UNP G0S8H7
A	-8	LYS	-	expression tag	UNP G0S8H7
A	-7	LEU	-	expression tag	UNP G0S8H7
A	-6	GLU	-	expression tag	UNP G0S8H7
A	-5	VAL	-	expression tag	UNP G0S8H7
A	-4	LEU	-	expression tag	UNP G0S8H7
A	-3	PHE	-	expression tag	UNP G0S8H7
A	-2	GLN	-	expression tag	UNP G0S8H7
A	-1	GLY	-	expression tag	UNP G0S8H7
A	0	PRO	-	expression tag	UNP G0S8H7

- Molecule 2 is a protein called Histone-lysine N-methyltransferase EZH2, Polycomb protein SUZ12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	794	Total	C	N	O	S	0	0	0
			6390	4022	1161	1166	41			

There are 17 discrepancies between the modelled and reference sequences:

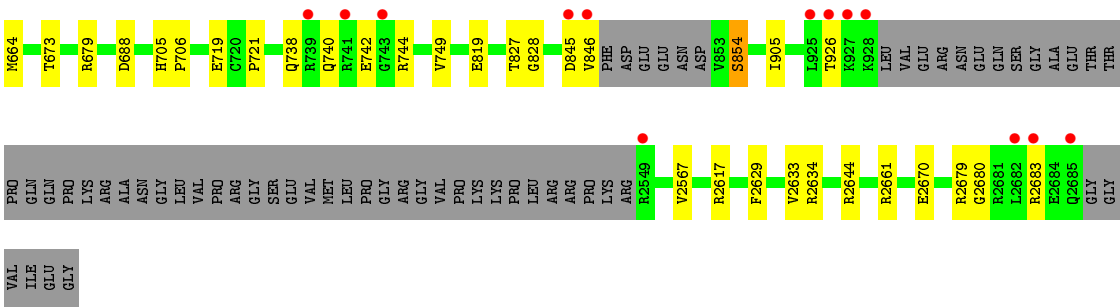
Chain	Residue	Modelled	Actual	Comment	Reference
B	182	SER	-	expression tag	UNP G0SDW4
B	183	ASN	-	expression tag	UNP G0SDW4
B	184	HIS	-	expression tag	UNP G0SDW4
B	185	HIS	-	expression tag	UNP G0SDW4
B	186	HIS	-	expression tag	UNP G0SDW4
B	187	HIS	-	expression tag	UNP G0SDW4
B	188	HIS	-	expression tag	UNP G0SDW4
B	189	HIS	-	expression tag	UNP G0SDW4
B	190	ALA	-	expression tag	UNP G0SDW4
B	840	ALA	GLU	engineered mutation	UNP G0SDW4
B	852	ASP	LYS	engineered mutation	UNP G0SDW4
B	2524	LEU	-	linker	UNP G0SDW4
B	2525	VAL	-	linker	UNP G0SDW4
B	2526	PRO	-	linker	UNP G0SDW4
B	2527	ARG	-	linker	UNP G0SDW4
B	2528	GLY	-	linker	UNP G0SDW4
B	2529	SER	-	linker	UNP G0SDW4

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	8	Total 8	Zn 8	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	351	Total 351	O 351	0	0
4	B	549	Total 549	O 549	0	0



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	117.59Å 138.00Å 224.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.75 – 2.11 44.75 – 2.11	Depositor EDS
% Data completeness (in resolution range)	94.8 (44.75-2.11) 94.8 (44.75-2.11)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.23 (at 2.12Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, R_{free}	0.168 , 0.214 0.168 , 0.214	Depositor DCC
R_{free} test set	4899 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	25.9	Xtriage
Anisotropy	0.071	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 60.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	10990	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.20% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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

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	0.42	0/3804	0.59	2/5176 (0.0%)
2	B	0.39	0/6537	0.54	0/8834
All	All	0.40	0/10341	0.56	2/14010 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	258	ARG	NE-CZ-NH1	5.76	123.18	120.30
1	A	258	ARG	NE-CZ-NH2	-5.33	117.64	120.30

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	3692	0	3565	25	0
2	B	6390	0	6290	62	0
3	B	8	0	0	0	0
4	A	351	0	0	3	0
4	B	549	0	0	9	0
All	All	10990	0	9855	86	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 4.

The worst 5 of 86 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)
2:B:660:LYS:O	2:B:744:ARG:NH1	1.92	1.01
2:B:2680:GLY:HA2	2:B:2683:ARG:HH11	1.26	0.99
2:B:579:GLY:O	2:B:630:ARG:NH1	2.11	0.83
2:B:2679:ARG:O	2:B:2683:ARG:HD2	1.79	0.82
2:B:2679:ARG:O	2:B:2683:ARG:CD	2.36	0.74

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	462/605 (76%)	443 (96%)	17 (4%)	2 (0%)	34	32
2	B	778/937 (83%)	745 (96%)	32 (4%)	1 (0%)	51	53
All	All	1240/1542 (80%)	1188 (96%)	49 (4%)	3 (0%)	47	48

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	393	ALA
1	A	558	LEU
2	B	411	SER

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	394/495 (80%)	389 (99%)	5 (1%)	69	74
2	B	695/815 (85%)	680 (98%)	15 (2%)	52	55
All	All	1089/1310 (83%)	1069 (98%)	20 (2%)	59	63

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	455	ARG
2	B	523	ARG
2	B	854	SER
2	B	276	TYR
2	B	367	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	397	GLN
2	B	313	ASN
2	B	609	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](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	468/605 (77%)	0.14	30 (6%) 19 23	12, 25, 79, 107	0
2	B	794/937 (84%)	0.08	62 (7%) 13 16	15, 34, 81, 134	0
All	All	1262/1542 (81%)	0.10	92 (7%) 15 19	12, 30, 80, 134	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	403	VAL	7.6
2	B	741	ARG	6.5
1	A	392	LEU	6.4
2	B	638	ARG	5.9
2	B	635	LEU	5.8

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	ZN	B	8008	1/1	0.94	0.04	55,55,55,55	0
3	ZN	B	8005	1/1	0.99	0.07	32,32,32,32	0
3	ZN	B	8004	1/1	0.99	0.06	32,32,32,32	0
3	ZN	B	8006	1/1	1.00	0.10	30,30,30,30	0
3	ZN	B	8007	1/1	1.00	0.11	18,18,18,18	0
3	ZN	B	8002	1/1	1.00	0.11	21,21,21,21	0
3	ZN	B	8003	1/1	1.00	0.10	21,21,21,21	0
3	ZN	B	8001	1/1	1.00	0.10	21,21,21,21	0

6.5 Other polymers [i](#)

There are no such residues in this entry.