



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 02:05 am BST

PDB ID : 3AV4
Title : Crystal structure of mouse DNA methyltransferase 1
Authors : Takeshita, K.; Suetake, I.; Yamashita, E.; Suga, M.; Narita, H.; Nakagawa, A.; Tajima, S.
Deposited on : 2011-02-22
Resolution : 2.75 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

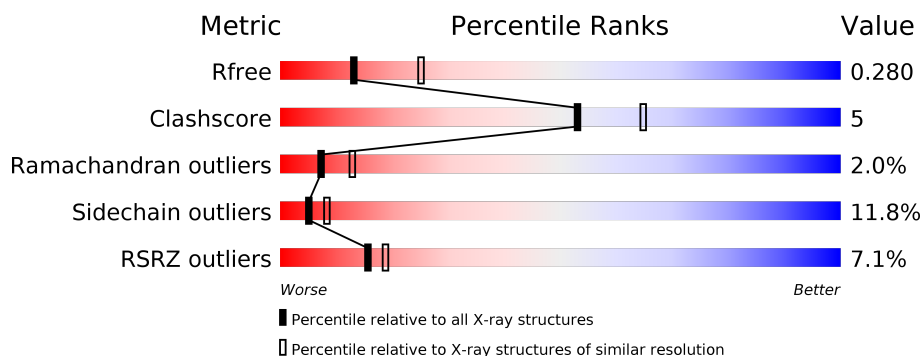
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	1330	<div> <div>6%</div> <div>65%</div> <div>18%</div> <div>•</div> <div>14%</div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9156 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 DNA (cytosine-5)-methyltransferase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1140	Total	C	N	O	S	0	0	0
			9110	5766	1586	1699	59			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	4	Total	Zn	0	0
			4	4		

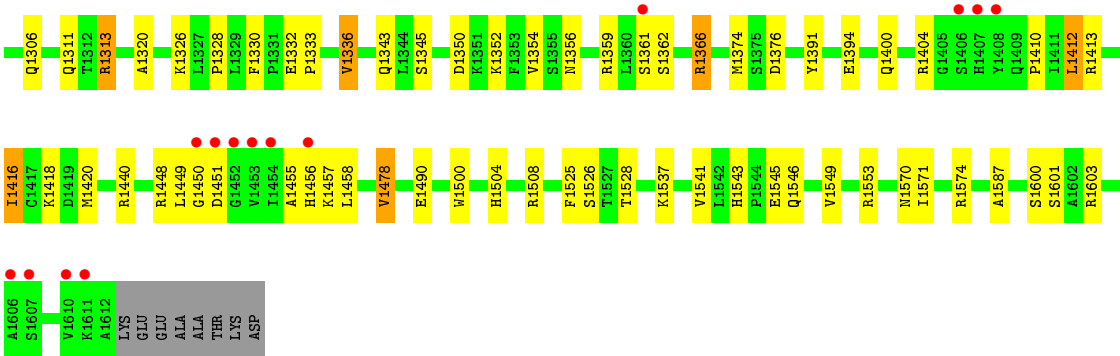
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	42	Total	O	0	0
			42	42		

These plots are drawn for all protein, RNA and DNA 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 ($\text{RSRZ} > 2$). 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.

- Chain A: 6% 65% 18% 14%

Residue	Amino Acid	Secondary ID
T1045	ALA	T1045
G1157	LYS	G1157
W1168	ARG	W1168
W1173	LYS	W1173
D1174	GLU	D1174
P1175	PRO	P1175
Q1178	GLN	Q1178
A1179	ALA	A1179
L1182	LEU	L1182
L1198	LEU	L1198
A1203	ALA	A1203
V1206	VAL	V1206
T1207	THR	T1207
G1211	GLY	G1211
G1212	GLY	G1212
R1213	ARG	R1213
D1219	ASP	D1219
Q1230	GLN	Q1230
G1231	GLY	G1231
F1232	PHE	F1232
S1233	SER	S1233
M1236	MET	M1236
R1237	ARG	R1237
R1241	ARG	R1241
K1245	LYS	K1245
F1266	PHE	F1266
M1270	MET	M1270
V1271	VAL	V1271
R1272	ARG	R1272
V1275	VAL	V1275
S1276	SER	S1276
I1277	ILE	I1277
R1278	ARG	R1278
R1279	ARG	R1279
R1288	ARG	R1288
G1289	GLY	G1289
R1292	ARG	R1292
M1293	MET	M1293
G1294	GLY	G1294
R1143	ARG	R1143
Y1026	TYR	Y1026



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	135.78Å 97.86Å 130.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.42 – 2.75 43.42 – 2.75	Depositor EDS
% Data completeness (in resolution range)	(Not available) (43.42-2.75) 99.7 (43.42-2.75)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.30 (at 2.77Å)	Xtriage
Refinement program	BUSTER-TNT, BUSTER 2.8.0	Depositor
R, R_{free}	0.230 , 0.267 0.237 , 0.280	Depositor DCC
R_{free} test set	2315 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	50.8	Xtriage
Anisotropy	0.075	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 57.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.013 for l,-k,h	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	9156	wwPDB-VP
Average B, all atoms (Å ²)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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	0.52	0/9330	0.79	3/12615 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	1490	GLU	C-N-CA	5.61	135.73	121.70
1	A	1042	ILE	C-N-CA	5.20	134.69	121.70
1	A	689	CYS	C-N-CA	5.16	134.59	121.70

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	9110	0	8892	98	0
2	A	4	0	0	0	0
3	A	42	0	0	0	0
All	All	9156	0	8892	98	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 5.

The worst 5 of 98 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:772:ASP:HB2	1:A:775:LYS:HG2	1.48	0.93
1:A:595:GLY:HA3	1:A:1504:HIS:CE1	2.19	0.78
1:A:1330:PHE:H	1:A:1356:ASN:HD21	1.35	0.75
1:A:975:GLU:OE1	1:A:978:ARG:HD2	1.87	0.75
1:A:479:GLU:HA	1:A:516:LYS:HE2	1.70	0.72

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	1118/1330 (84%)	1007 (90%)	89 (8%)	22 (2%)	7 13

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	380	ALA
1	A	456	GLU
1	A	465	GLY
1	A	573	ALA
1	A	603	ALA

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	1002/1162 (86%)	884 (88%)	118 (12%)	5 8

5 of 118 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	885	LYS
1	A	989	LEU
1	A	1458	LEU
1	A	898	LEU
1	A	911	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1098	ASN
1	A	1159	HIS
1	A	1407	HIS
1	A	917	GLN
1	A	1430	HIS

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 4 ligands modelled in this entry, 4 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

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 [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	1140/1330 (85%)	0.41	81 (7%) 16 19	21, 59, 104, 128	0

The worst 5 of 81 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1072	LEU	6.4
1	A	1406	SER	6.3
1	A	390	GLU	5.5
1	A	489	PHE	5.2
1	A	1012	VAL	5.1

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(Å ²)	Q<0.9
2	ZN	A	2001	1/1	0.98	0.06	35,35,35,35	0
2	ZN	A	2002	1/1	0.98	0.02	36,36,36,36	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	ZN	A	2004	1/1	0.99	0.03	32,32,32,32	0
2	ZN	A	2005	1/1	0.99	0.05	9,9,9,9	0

6.5 Other polymers [i](#)

There are no such residues in this entry.