



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2016 – 12:50 PM EST

PDB ID : 5HEK
Title : crystal structure of M1.HpyAVI
Authors : Ma, B.; Zhang, H.; Liu, W.
Deposited on : 2016-01-06
Resolution : 3.00 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

<http://wwpdb.org/validation/2016/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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028320

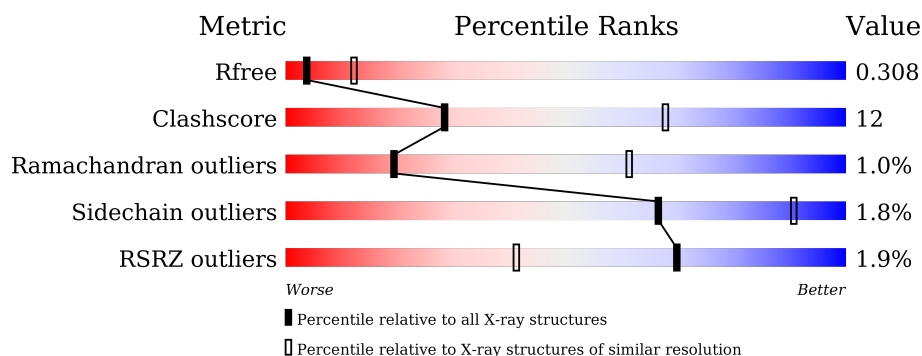
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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. 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	240	<div> <div>%</div> <div> <div></div> <div>62%</div> <div>18%</div> <div>•</div> <div>18%</div> </div> </div>
1	B	240	<div> <div>2%</div> <div> <div></div> <div>54%</div> <div>22%</div> <div>•</div> <div>23%</div> </div> </div>
1	C	240	<div> <div>2%</div> <div> <div></div> <div>57%</div> <div>19%</div> <div>•</div> <div>23%</div> </div> </div>
1	D	240	<div> <div>%</div> <div> <div></div> <div>54%</div> <div>18%</div> <div>•</div> <div>26%</div> </div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6166 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 Adenine specific DNA methyltransferase (DpnA).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	197	Total	C	N	O	S	0	0	0
			1638	1074	274	282	8			
1	B	185	Total	C	N	O	S	0	0	0
			1532	1004	254	266	8			
1	C	185	Total	C	N	O	S	0	0	0
			1533	1004	257	265	7			
1	D	177	Total	C	N	O	S	0	0	0
			1463	962	243	251	7			

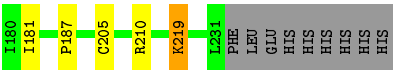
There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	233	LEU	-	expression tag	UNP O24891
A	234	GLU	-	expression tag	UNP O24891
A	235	HIS	-	expression tag	UNP O24891
A	236	HIS	-	expression tag	UNP O24891
A	237	HIS	-	expression tag	UNP O24891
A	238	HIS	-	expression tag	UNP O24891
A	239	HIS	-	expression tag	UNP O24891
A	240	HIS	-	expression tag	UNP O24891
B	233	LEU	-	expression tag	UNP O24891
B	234	GLU	-	expression tag	UNP O24891
B	235	HIS	-	expression tag	UNP O24891
B	236	HIS	-	expression tag	UNP O24891
B	237	HIS	-	expression tag	UNP O24891
B	238	HIS	-	expression tag	UNP O24891
B	239	HIS	-	expression tag	UNP O24891
B	240	HIS	-	expression tag	UNP O24891
C	233	LEU	-	expression tag	UNP O24891
C	234	GLU	-	expression tag	UNP O24891
C	235	HIS	-	expression tag	UNP O24891
C	236	HIS	-	expression tag	UNP O24891
C	237	HIS	-	expression tag	UNP O24891

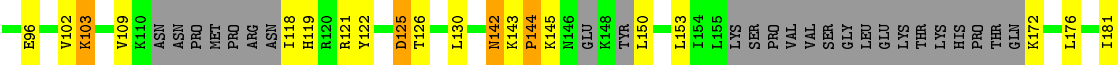
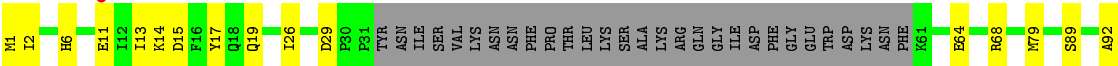
Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	238	HIS	-	expression tag	UNP O24891
C	239	HIS	-	expression tag	UNP O24891
C	240	HIS	-	expression tag	UNP O24891
D	233	LEU	-	expression tag	UNP O24891
D	234	GLU	-	expression tag	UNP O24891
D	235	HIS	-	expression tag	UNP O24891
D	236	HIS	-	expression tag	UNP O24891
D	237	HIS	-	expression tag	UNP O24891
D	238	HIS	-	expression tag	UNP O24891
D	239	HIS	-	expression tag	UNP O24891
D	240	HIS	-	expression tag	UNP O24891



● Molecule 1: Adenine specific DNA methyltransferase (DpnA)



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	69.72Å 69.72Å 532.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.09 – 3.00 48.16 – 3.00	Depositor EDS
% Data completeness (in resolution range)	97.6 (49.09-3.00) 99.0 (48.16-3.00)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.67 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.251 , 0.308 0.252 , 0.308	Depositor DCC
R_{free} test set	1255 reflections (4.61%)	DCC
Wilson B-factor (Å ²)	55.0	Xtriage
Anisotropy	0.489	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 46.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.69	EDS
Total number of atoms	6166	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 9.11% 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

5.1 Standard geometry

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	0.41	0/1680	0.72	2/2268 (0.1%)
1	B	0.40	0/1569	0.62	0/2117
1	C	0.39	0/1570	0.67	0/2118
1	D	0.40	0/1495	0.64	0/2012
All	All	0.40	0/6314	0.66	2/8515 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	4

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	114	MET	C-N-CD	-6.75	105.76	120.60
1	A	155	LEU	CA-CB-CG	6.61	130.49	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	142	ASN	Peptide
1	A	145	LYS	Peptide
1	A	147	GLU	Peptide
1	A	60	PHE	Peptide

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	1638	0	1662	43	0
1	B	1532	0	1544	45	0
1	C	1533	0	1559	44	0
1	D	1463	0	1496	34	0
All	All	6166	0	6261	144	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 12.

The worst 5 of 144 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:229:LEU:O	1:B:230:ASN:ND2	2.14	0.81
1:C:146:ASN:OD1	1:C:147:GLU:N	2.14	0.79
1:B:29:ASP:HB2	1:B:82:PHE:HD2	1.46	0.79
1:B:119:HIS:HA	1:D:143:LYS:HE3	1.66	0.76
1:C:6:HIS:CE1	1:C:219:LYS:HD2	2.21	0.76

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	191/240 (80%)	169 (88%)	20 (10%)	2 (1%)	19	61
1	B	176/240 (73%)	155 (88%)	20 (11%)	1 (1%)	30	72

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	177/240 (74%)	159 (90%)	15 (8%)	3 (2%)	11	46
1	D	166/240 (69%)	152 (92%)	13 (8%)	1 (1%)	30	72
All	All	710/960 (74%)	635 (89%)	68 (10%)	7 (1%)	19	61

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	149	TYR
1	A	155	LEU
1	C	144	PRO
1	D	144	PRO
1	B	144	PRO

5.3.2 Protein sidechains ⓘ

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	179/218 (82%)	178 (99%)	1 (1%)	90	97
1	B	167/218 (77%)	164 (98%)	3 (2%)	66	91
1	C	167/218 (77%)	165 (99%)	2 (1%)	78	94
1	D	159/218 (73%)	153 (96%)	6 (4%)	40	78
All	All	672/872 (77%)	660 (98%)	12 (2%)	66	91

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	219	LYS
1	D	11	GLU
1	D	142	ASN
1	C	138	LYS
1	D	125	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	6	HIS
1	C	112	ASN
1	C	142	ASN
1	C	186	ASN
1	D	142	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 carbohydrates 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	197/240 (82%)	-0.08	3 (1%) 76 49	33, 56, 99, 142	0
1	B	185/240 (77%)	-0.05	5 (2%) 58 28	33, 56, 93, 140	0
1	C	185/240 (77%)	0.02	4 (2%) 65 35	37, 61, 103, 145	0
1	D	177/240 (73%)	-0.12	2 (1%) 82 58	36, 59, 90, 108	0
All	All	744/960 (77%)	-0.06	14 (1%) 70 41	33, 58, 99, 145	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	33	ASN	3.2
1	C	117	ASN	2.9
1	A	146	ASN	2.6
1	B	220	GLU	2.4
1	C	116	ARG	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 carbohydrates 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.