



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2017 – 11:19 PM EDT

PDB ID : 3DP7
Title : CRYSTAL STRUCTURE OF SAM-dependent methyltransferase from *Bacteroides vulgatus* ATCC 8482
Authors : Malashkevich, V.N.; Toro, R.; Ramagopal, U.; Meyer, A.J.; Sauder, J.M.; Bursley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : unknown
Resolution : 2.33 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

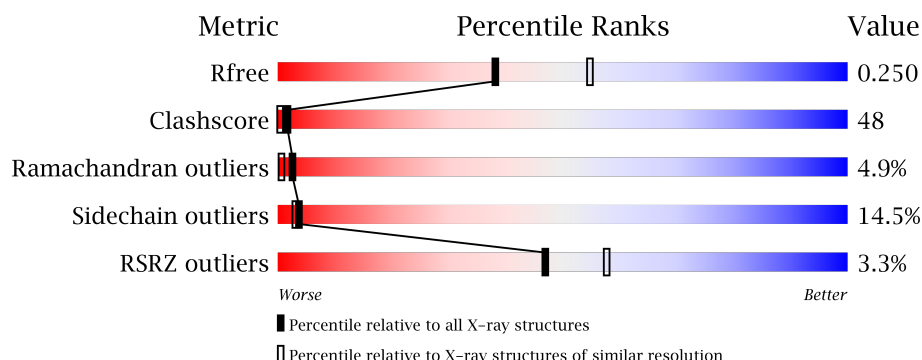
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.33 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}	100719	1570 (2.36-2.32)
Clashscore	112137	1673 (2.36-2.32)
Ramachandran outliers	110173	1654 (2.36-2.32)
Sidechain outliers	110143	1655 (2.36-2.32)
RSRZ outliers	101464	1576 (2.36-2.32)

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	363	<div> <div>4%</div> <div> <div></div> <div>44%</div> <div>39%</div> <div>12%</div> <div>• •</div> </div> </div>
1	B	363	<div> <div>2%</div> <div> <div></div> <div>50%</div> <div>36%</div> <div>9%</div> <div>• •</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5694 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 SAM-dependent methyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	351	Total	C	N	O	S	0	0	0
			2808	1789	477	527	15			
1	B	350	Total	C	N	O	S	0	0	0
			2796	1780	476	525	15			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP A6KZ44
A	2	SER	-	expression tag	UNP A6KZ44
A	3	LEU	-	expression tag	UNP A6KZ44
A	356	GLU	-	expression tag	UNP A6KZ44
A	357	GLY	-	expression tag	UNP A6KZ44
A	358	HIS	-	expression tag	UNP A6KZ44
A	359	HIS	-	expression tag	UNP A6KZ44
A	360	HIS	-	expression tag	UNP A6KZ44
A	361	HIS	-	expression tag	UNP A6KZ44
A	362	HIS	-	expression tag	UNP A6KZ44
A	363	HIS	-	expression tag	UNP A6KZ44
B	1	MET	-	expression tag	UNP A6KZ44
B	2	SER	-	expression tag	UNP A6KZ44
B	3	LEU	-	expression tag	UNP A6KZ44
B	356	GLU	-	expression tag	UNP A6KZ44
B	357	GLY	-	expression tag	UNP A6KZ44
B	358	HIS	-	expression tag	UNP A6KZ44
B	359	HIS	-	expression tag	UNP A6KZ44
B	360	HIS	-	expression tag	UNP A6KZ44
B	361	HIS	-	expression tag	UNP A6KZ44
B	362	HIS	-	expression tag	UNP A6KZ44
B	363	HIS	-	expression tag	UNP A6KZ44

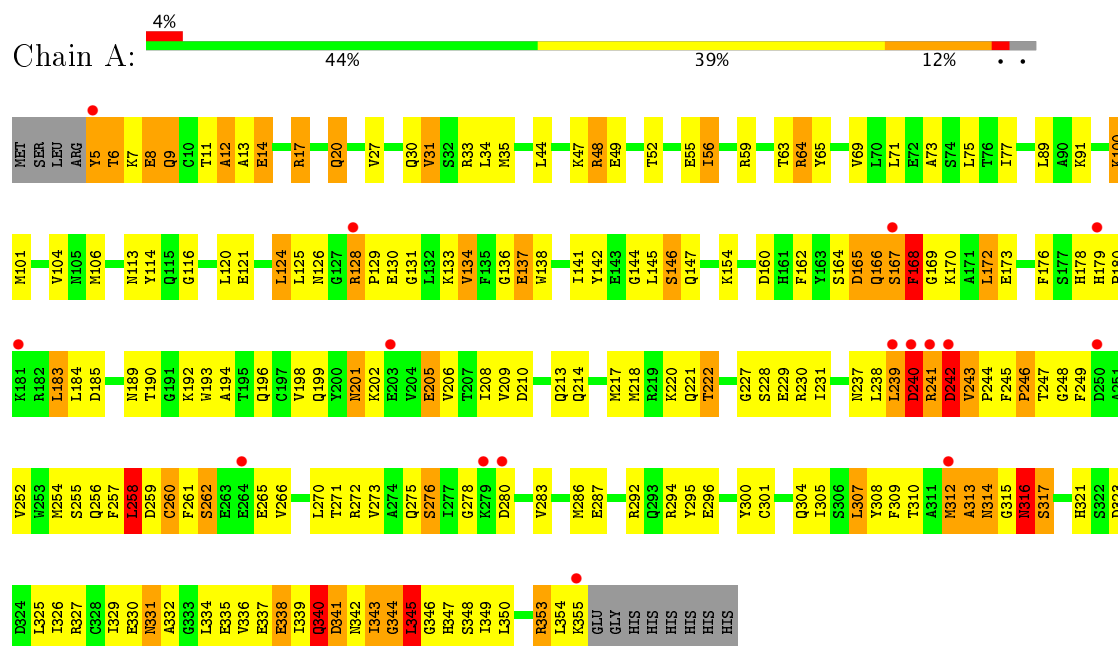
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	41	Total 41	O 41	0	0
2	B	49	Total 49	O 49	0	0

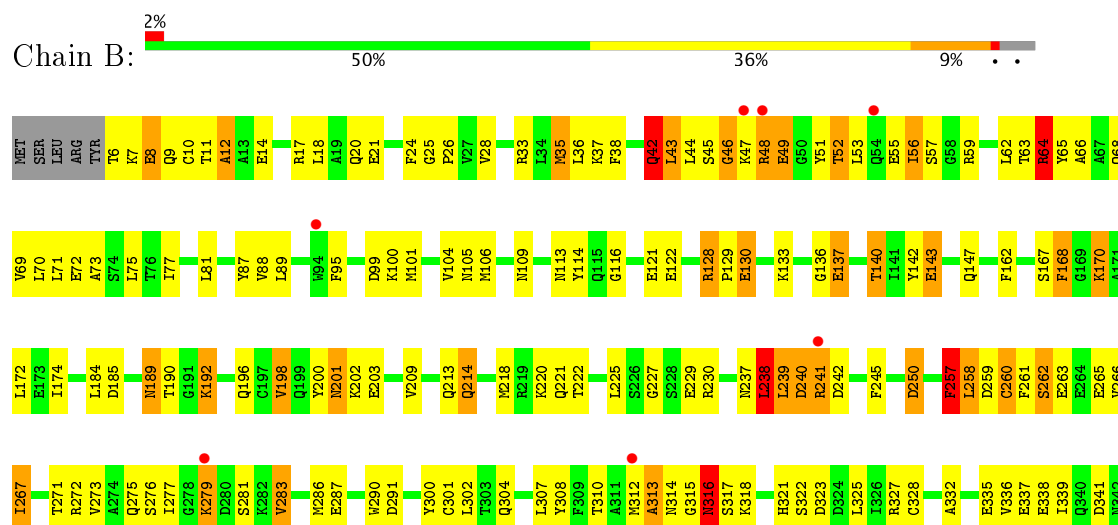
3 Residue-property plots [i](#)

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.

• Molecule 1: SAM-dependent methyltransferase



• Molecule 1: SAM-dependent methyltransferase



I343	I344	I345	I346	I347	I348	I349	I350	I351	I352	I353	I354	I355	GLU	GLY	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS
------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

4 Data and refinement statistics

Property	Value	Source
Space group	P 64	Depositor
Cell constants a, b, c, α , β , γ	130.80Å 130.80Å 122.75Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.82 – 2.33 19.82 – 2.33	Depositor EDS
% Data completeness (in resolution range)	65.0 (19.82-2.33) 65.1 (19.82-2.33)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.08 (at 2.33Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.183 , 0.253 0.182 , 0.250	Depositor DCC
R_{free} test set	1671 reflections (5.34%)	DCC
Wilson B-factor (Å ²)	60.6	Xtriage
Anisotropy	0.073	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 45.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.114 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5694	wwPDB-VP
Average B, all atoms (Å ²)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.51% 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	0.57	0/2867	0.75	6/3872 (0.2%)
1	B	0.56	0/2854	0.73	2/3854 (0.1%)
All	All	0.57	0/5721	0.74	8/7726 (0.1%)

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	1

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	81	LEU	CA-CB-CG	7.33	132.16	115.30
1	A	258	LEU	CA-CB-CG	5.78	128.60	115.30
1	B	257	PHE	N-CA-C	-5.66	95.71	111.00
1	A	317	SER	N-CA-C	-5.45	96.28	111.00
1	A	316	ASN	N-CA-C	-5.42	96.37	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	167	SER	Peptide

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	2808	0	2755	323	0
1	B	2796	0	2745	220	0
2	A	41	0	0	1	0
2	B	49	0	0	5	0
All	All	5694	0	5500	531	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 48.

The worst 5 of 531 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:316:ASN:HB2	1:A:317:SER:CB	1.38	1.53
1:B:52:THR:CG2	1:B:55:GLU:HG3	1.39	1.50
1:B:33:ARG:HA	1:B:106:MET:CE	1.40	1.49
1:A:316:ASN:CB	1:A:317:SER:HB2	1.49	1.42
1:B:314:ASN:HB2	1:B:316:ASN:ND2	1.40	1.37

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	349/363 (96%)	294 (84%)	33 (10%)	22 (6%)	1	0
1	B	348/363 (96%)	315 (90%)	21 (6%)	12 (3%)	4	2

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	697/726 (96%)	609 (87%)	54 (8%)	34 (5%)	2 0

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	6	THR
1	A	8	GLU
1	A	12	ALA
1	A	165	ASP
1	A	166	GLN

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	301/312 (96%)	259 (86%)	42 (14%)	4 3
1	B	300/312 (96%)	255 (85%)	45 (15%)	3 3
All	All	601/624 (96%)	514 (86%)	87 (14%)	4 3

5 of 87 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	343	ILE
1	B	49	GLU
1	B	308	TYR
1	A	345	LEU
1	B	35	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	331	ASN
1	B	20	GLN
1	B	321	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	340	GLN
1	B	9	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](#)

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	351/363 (96%)	0.17	16 (4%) 33 44	43, 62, 84, 104	0
1	B	350/363 (96%)	0.06	7 (2%) 65 74	46, 63, 80, 94	0
All	All	701/726 (96%)	0.12	23 (3%) 47 58	43, 62, 83, 104	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	242	ASP	5.9
1	A	240	ASP	4.4
1	A	241	ARG	4.4
1	A	181	LYS	4.1
1	A	279	LYS	4.1

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.