



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 30, 2017 – 09:10 PM EDT

PDB ID : 3ME5  
Title : Crystal structure of putative dna cytosine methylase from shigella flexneri 2a str. 2457T  
Authors : Ramagopal, U.A.; Malashkevich, V.N.; Toro, R.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYS-GXRC)  
Deposited on : unknown  
Resolution : 1.75 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](mailto: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

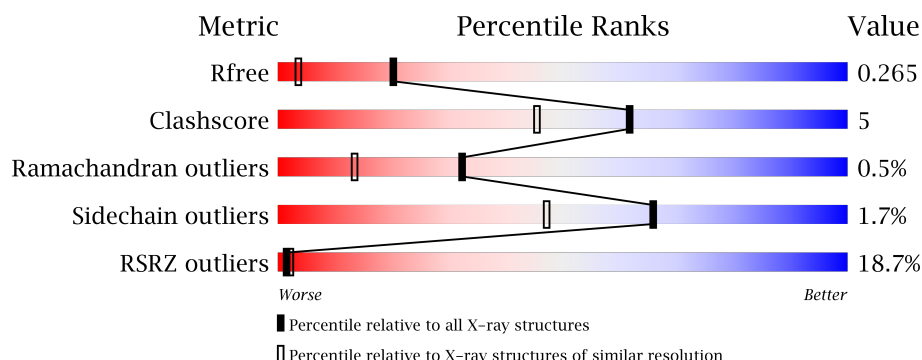
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.75 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))
$R_{free}$	100719	1762 (1.76-1.76)
Clashscore	112137	1889 (1.76-1.76)
Ramachandran outliers	110173	1868 (1.76-1.76)
Sidechain outliers	110143	1868 (1.76-1.76)
RSRZ outliers	101464	1770 (1.76-1.76)

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  $\geq 3$ , 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	482	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3628 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 Cytosine-specific methyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	414	Total	C	N	O	S	Se	0	4	0
			3350	2129	602	605	6	8			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MSE	-	EXPRESSION TAG	UNP E3Y0J2
A	0	SER	-	EXPRESSION TAG	UNP E3Y0J2
A	1	LEU	-	EXPRESSION TAG	UNP E3Y0J2
A	473	GLU	-	EXPRESSION TAG	UNP E3Y0J2
A	474	GLY	-	EXPRESSION TAG	UNP E3Y0J2
A	475	HIS	-	EXPRESSION TAG	UNP E3Y0J2
A	476	HIS	-	EXPRESSION TAG	UNP E3Y0J2
A	477	HIS	-	EXPRESSION TAG	UNP E3Y0J2
A	478	HIS	-	EXPRESSION TAG	UNP E3Y0J2
A	479	HIS	-	EXPRESSION TAG	UNP E3Y0J2
A	480	HIS	-	EXPRESSION TAG	UNP E3Y0J2

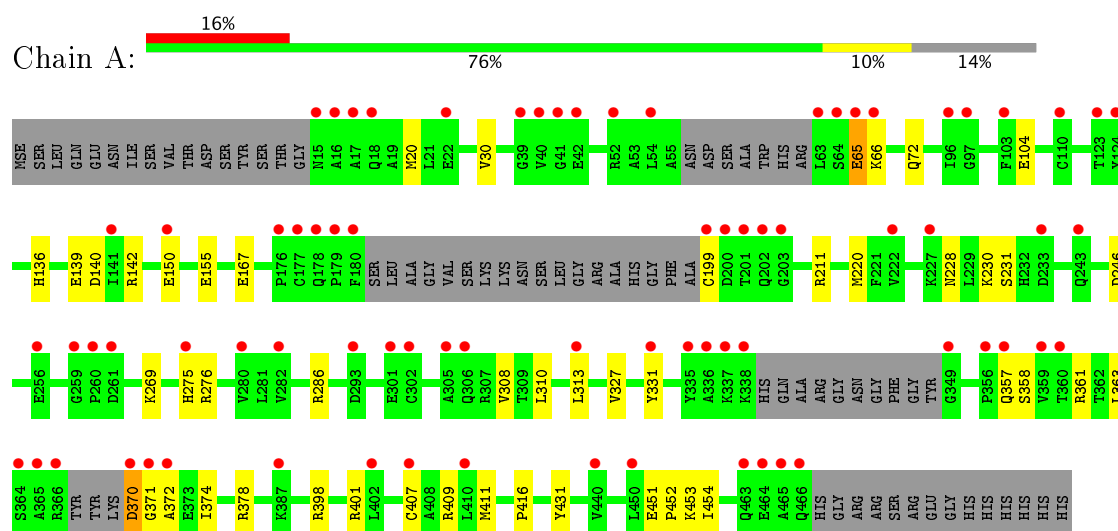
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	278	Total	O	0	0
			278	278		

### 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 ( $RSRZ > 2$ ). 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: Cytosine-specific methyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	149.63 Å 51.59 Å 63.03 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.75 20.00 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.1 (20.00-1.75) 99.1 (20.00-1.75)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.37 (at 1.76 Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.220 , 0.261 0.225 , 0.265	Depositor DCC
$R_{free}$ test set	2520 reflections (5.36%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.3	Xtriage
Anisotropy	0.221	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 41.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3628	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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](#)

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	0/3426	0.64	0/4620

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	3350	0	3309	34	0
2	A	278	0	0	9	0
All	All	3628	0	3309	34	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.

All (34) 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:308:VAL:HG11	1:A:313:LEU:HD21	1.69	0.74
1:A:308:VAL:CG1	1:A:313:LEU:HD21	2.19	0.72
1:A:150:GLU:H	1:A:150:GLU:CD	1.95	0.69
1:A:371:GLY:HA2	1:A:401:ARG:HB3	1.74	0.68

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:370:ASP:HA	1:A:374:ILE:HB	1.76	0.68
1:A:370:ASP:HB2	1:A:407:CYS:SG	2.39	0.63
1:A:371:GLY:HA3	1:A:431:TYR:CZ	2.36	0.61
1:A:199:CYS:N	2:A:677:HOH:O	2.35	0.60
1:A:269:LYS:HD3	1:A:275:HIS:CE1	2.39	0.56
1:A:139:GLU:HG2	2:A:491:HOH:O	2.05	0.56
1:A:230:LYS:HB2	2:A:633:HOH:O	2.06	0.55
1:A:220:MSE:HE1	1:A:454:ILE:HA	1.89	0.55
1:A:370:ASP:N	1:A:374:ILE:H	2.04	0.54
1:A:276:ARG:NE	1:A:411:MSE:HE1	2.23	0.53
1:A:136[A]:HIS:ND1	2:A:591:HOH:O	2.34	0.51
1:A:140:ASP:HB3	2:A:756:HOH:O	2.12	0.49
1:A:104:GLU:CG	2:A:635:HOH:O	2.63	0.47
1:A:228:ASN:ND2	1:A:228:ASN:H	2.13	0.46
1:A:327:VAL:HG12	1:A:331:TYR:CE2	2.50	0.46
1:A:246:ASP:OD1	1:A:286:ARG:NH2	2.48	0.46
1:A:371:GLY:HA3	1:A:431:TYR:CE1	2.51	0.45
1:A:451:GLU:HB3	1:A:452:PRO:HD3	1.99	0.45
1:A:276:ARG:HE	1:A:411:MSE:HE1	1.80	0.45
1:A:155[B]:GLU:OE1	1:A:155[B]:GLU:HA	2.16	0.44
1:A:142:ARG:HG2	2:A:756:HOH:O	2.18	0.44
1:A:30:VAL:HG23	2:A:625:HOH:O	2.18	0.43
1:A:228:ASN:H	1:A:228:ASN:HD22	1.65	0.43
1:A:104:GLU:HG2	2:A:635:HOH:O	2.20	0.42
1:A:65:GLU:HB3	1:A:66:LYS:H	1.71	0.42
1:A:358:SER:HA	1:A:361:ARG:HG2	2.01	0.42
1:A:378:ARG:HD3	1:A:398:ARG:O	2.20	0.41
1:A:409:ARG:HD3	1:A:416:PRO:HA	2.01	0.41
1:A:310:LEU:HD11	1:A:363:LEU:HD11	2.03	0.41
1:A:20:MSE:HE1	1:A:72:GLN:CG	2.50	0.40

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	408/482 (85%)	396 (97%)	10 (2%)	2 (0%)	32	13

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	65	GLU
1	A	372	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	355/396 (90%)	348 (98%)	7 (2%)	60	38

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	167[A]	GLU
1	A	167[B]	GLU
1	A	211	ARG
1	A	231	SER
1	A	357	GLN
1	A	370	ASP
1	A	453	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	228	ASN
1	A	258	ASN
1	A	357	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	406/482 (84%)	1.14	76 (18%) ⓘ ⓘ	14, 24, 55, 68	0

All (76) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	359	VAL	9.1
1	A	15	ASN	7.5
1	A	360	THR	7.4
1	A	199	CYS	7.1
1	A	366	ARG	6.7
1	A	52	ARG	5.8
1	A	338	LYS	5.8
1	A	371	GLY	5.6
1	A	356	PRO	5.6
1	A	466	GLN	5.3
1	A	16	ALA	5.3
1	A	365	ALA	4.9
1	A	200	ASP	4.8
1	A	335	TYR	4.7
1	A	179	PRO	4.5
1	A	337	LYS	4.3
1	A	180	PHE	4.1
1	A	41	GLY	4.0
1	A	150	GLU	4.0
1	A	202	GLN	3.8
1	A	17	ALA	3.7
1	A	227	LYS	3.7
1	A	464	GLU	3.7
1	A	54	LEU	3.7
1	A	465	ALA	3.6
1	A	370	ASP	3.5
1	A	357	GLN	3.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	463	GLN	3.2
1	A	336	ALA	3.2
1	A	233	ASP	3.2
1	A	364	SER	3.2
1	A	305	ALA	3.0
1	A	65	GLU	3.0
1	A	260	PRO	3.0
1	A	256	GLU	3.0
1	A	96	ILE	2.9
1	A	302	CYS	2.9
1	A	63	LEU	2.9
1	A	402	LEU	2.9
1	A	18	GLN	2.8
1	A	306	GLN	2.8
1	A	177	CYS	2.8
1	A	349	GLY	2.8
1	A	176	PRO	2.8
1	A	64	SER	2.7
1	A	203	GLY	2.7
1	A	410	LEU	2.7
1	A	261	ASP	2.7
1	A	39	GLY	2.5
1	A	293	ASP	2.5
1	A	407	CYS	2.4
1	A	372	ALA	2.4
1	A	222	VAL	2.4
1	A	259	GLY	2.4
1	A	97	GLY	2.3
1	A	275	HIS	2.3
1	A	42	GLU	2.3
1	A	440	VAL	2.3
1	A	124	TYR	2.3
1	A	450	LEU	2.3
1	A	40	VAL	2.2
1	A	201	THR	2.2
1	A	178	GLN	2.2
1	A	280	VAL	2.2
1	A	282	VAL	2.2
1	A	313	LEU	2.2
1	A	141	ILE	2.2
1	A	301	GLU	2.1
1	A	123	THR	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	66	LYS	2.1
1	A	387	LYS	2.1
1	A	110	CYS	2.1
1	A	331	TYR	2.0
1	A	103	PHE	2.0
1	A	22	GLU	2.0
1	A	243	GLN	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 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.