



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 21, 2016 – 04:45 PM EDT

PDB ID : 5DOO  
Title : The structure of PKMT2 from Rickettsia typhi  
Authors : Noinaj, N.; Abeykoon, A.; He, Y.; Yang, D.C.; Buchanan, S.K.  
Deposited on : 2015-09-11  
Resolution : 3.13 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](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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
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-20027939

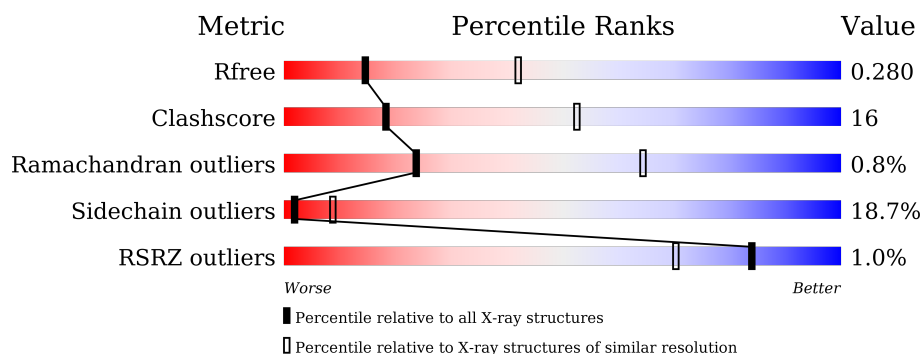
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.13 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	1095 (3.18-3.10)
Clashscore	102246	1202 (3.18-3.10)
Ramachandran outliers	100387	1162 (3.18-3.10)
Sidechain outliers	100360	1162 (3.18-3.10)
RSRZ outliers	91569	1097 (3.18-3.10)

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  $\geq 3$ , 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	535	<div> <div>%</div> <div> <div></div> <div>54%</div> <div>34%</div> <div>7%</div> <div>5%</div> </div> </div>
1	B	535	<div> <div>%</div> <div> <div></div> <div>53%</div> <div>34%</div> <div>9%</div> <div>.</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8167 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 protein lysine methyltransferase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	510	Total	C	N	O	S	0	0	0
			4072	2638	660	760	14			
1	B	513	Total	C	N	O	S	0	0	0
			4093	2652	666	760	15			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q68XQ5
B	0	GLY	-	expression tag	UNP Q68XQ5

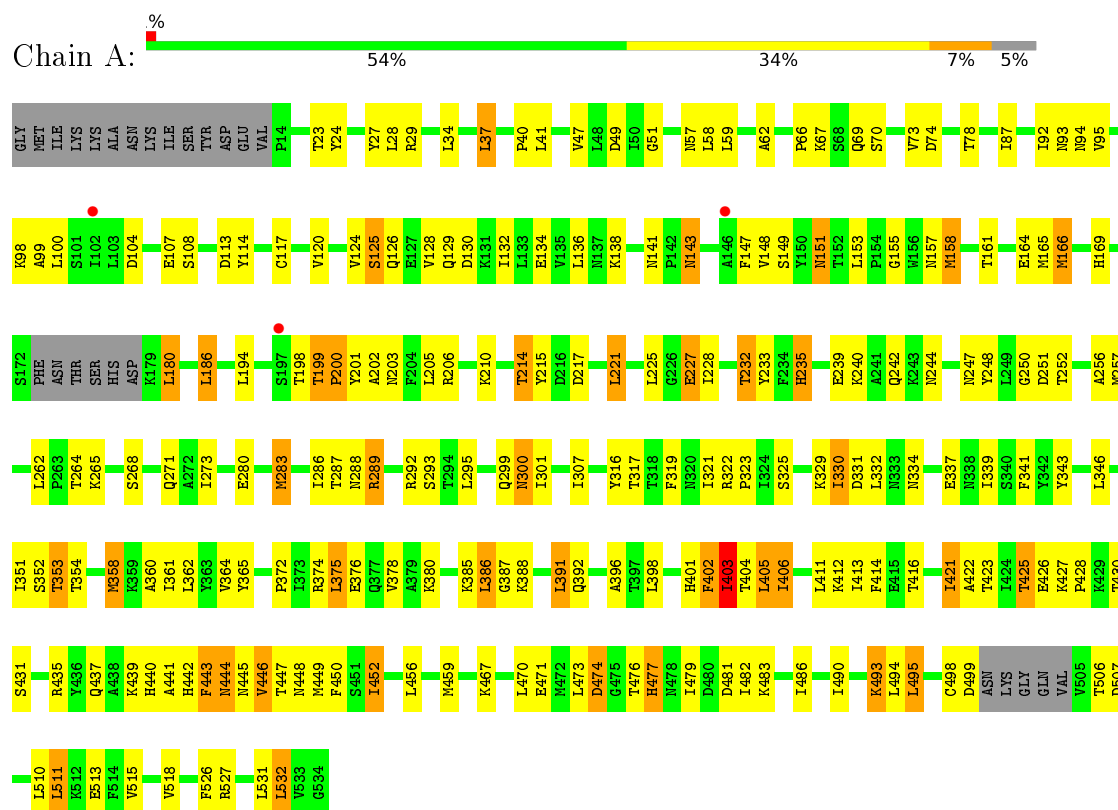
- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Ca	0	0
			2	2		

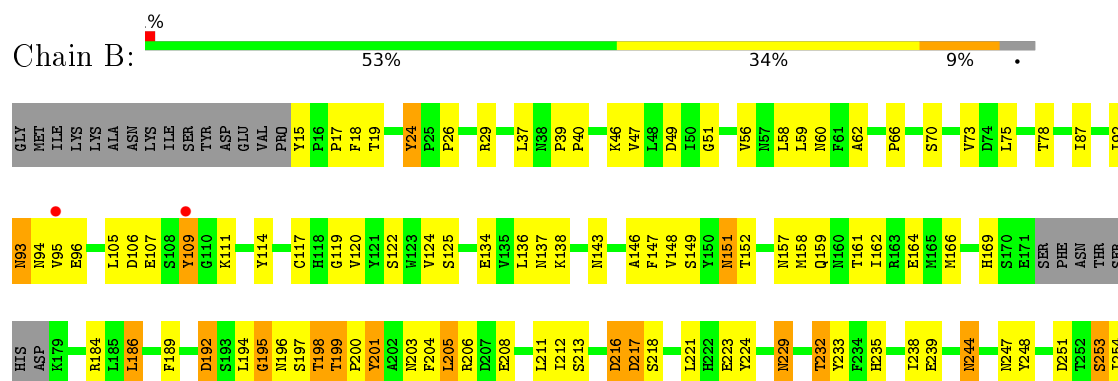
### 3 Residue-property plots

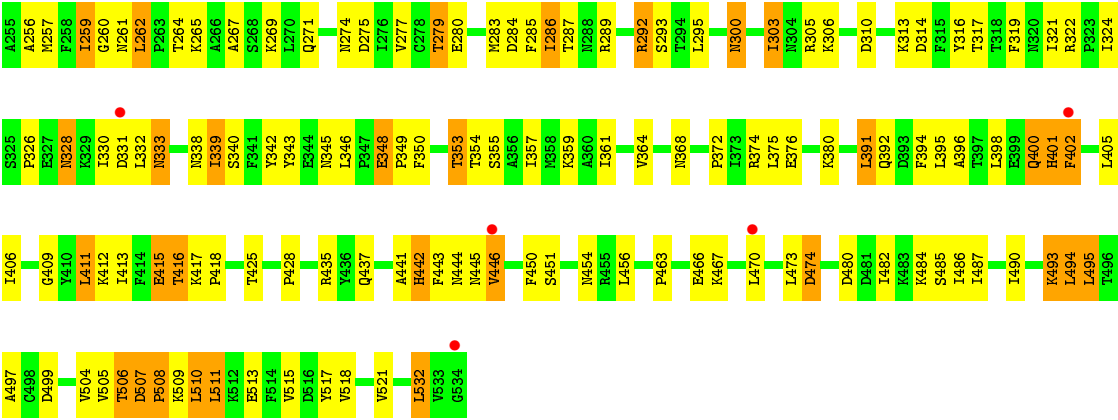
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: protein lysine methyltransferase 2



#### • Molecule 1: protein lysine methyltransferase 2





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.72Å 90.53Å 106.20Å 90.00° 114.17° 90.00°	Depositor
Resolution (Å)	48.45 – 3.13 48.45 – 3.13	Depositor EDS
% Data completeness (in resolution range)	96.4 (48.45-3.13) 96.2 (48.45-3.13)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.96 (at 3.12Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.233 , 0.280 0.231 , 0.280	Depositor DCC
$R_{free}$ test set	1243 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.4	Xtriage
Anisotropy	0.242	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 50.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.055 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	8167	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.38% 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

### 5.1 Standard geometry

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

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	0.31	0/4162	0.55	2/5646 (0.0%)
1	B	0.31	0/4184	0.56	0/5675
All	All	0.31	0/8346	0.55	2/11321 (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	1
1	B	0	2
All	All	0	3

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	200	PRO	C-N-CA	8.58	143.15	121.70
1	A	386	LEU	CA-CB-CG	5.15	127.14	115.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	227	GLU	Peptide
1	B	195	GLY	Peptide
1	B	507	ASP	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	4072	0	4044	123	0
1	B	4093	0	4066	133	0
2	A	2	0	0	0	0
All	All	8167	0	8110	254	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 16.

The worst 5 of 254 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:452:ILE:O	1:A:459:MET:HA	1.64	0.96
1:A:227:GLU:HB3	1:A:228:ILE:HA	1.61	0.83
1:A:217:ASP:O	1:A:221:LEU:HB2	1.78	0.82
1:A:435:ARG:NH1	1:A:473:LEU:O	2.15	0.80
1:B:396:ALA:O	1:B:400:GLN:HB2	1.81	0.80

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	504/535 (94%)	455 (90%)	47 (9%)	2 (0%)	39 77
1	B	509/535 (95%)	458 (90%)	45 (9%)	6 (1%)	16 53
All	All	1013/1070 (95%)	913 (90%)	92 (9%)	8 (1%)	24 64

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	505	VAL
1	B	446	VAL
1	B	196	ASN
1	B	198	THR
1	B	200	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	447/484 (92%)	363 (81%)	84 (19%)	2	9
1	B	447/484 (92%)	364 (81%)	83 (19%)	2	9
All	All	894/968 (92%)	727 (81%)	167 (19%)	2	9

5 of 167 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	495	LEU
1	B	96	GLU
1	B	470	LEU
1	A	499	ASP
1	B	15	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	129	GLN
1	B	196	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 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, 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	510/535 (95%)	-0.05	3 (0%) 90 81	35, 70, 113, 176	0
1	B	513/535 (95%)	0.02	7 (1%) 78 61	41, 79, 133, 185	0
All	All	1023/1070 (95%)	-0.01	10 (0%) 84 70	35, 74, 126, 185	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	SER	3.4
1	B	331	ASP	2.9
1	B	534	GLY	2.8
1	B	109	TYR	2.6
1	B	446	VAL	2.5

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	CA	A	602	1/1	0.96	0.12	-1.92	87,87,87,87	0
2	CA	A	601	1/1	0.92	0.21	-	55,55,55,55	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.