



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 9, 2020 – 12:49 AM EST

PDB ID : 7KWJ  
Title : Spermidine N-acetyltransferase SpeG K23-Q34 chimera from *Vibrio cholerae* and hSSAT  
Authors : Le, V.T.B.; Tsimbalyuk, S.; Lim, E.Q.; Solis, A.; Gawat, D.; Boeck, P.; Renolo, R.; Forwood, J.K.; Kuhn, M.L.  
Deposited on : 2020-12-01  
Resolution : 2.58 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

<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	:	2.15.1
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)	:	2.15.1

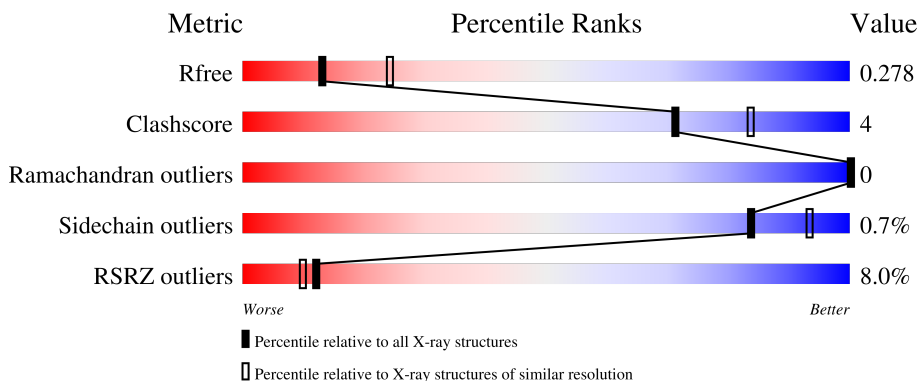
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.58 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	3676 (2.60-2.56)
Clashscore	141614	4049 (2.60-2.56)
Ramachandran outliers	138981	3979 (2.60-2.56)
Sidechain outliers	138945	3979 (2.60-2.56)
RSRZ outliers	127900	3614 (2.60-2.56)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	173	 7% 87% 8% 6%
1	B	173	 8% 89% 7% .
1	C	173	 8% 81% 11% . 8%

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 8005 atoms, of which 3939 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 Spermidine N(1)-acetyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	163	Total	C	H	N	O	S	0	0	0
			2660	865	1308	237	248	2			
1	B	166	Total	C	H	N	O	S	0	0	0
			2711	881	1335	240	253	2			
1	C	160	Total	C	H	N	O	S	0	0	0
			2634	858	1296	233	245	2			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	LYS	ASN	engineered mutation	UNP Q9KL03
A	24	GLU	ASN	engineered mutation	UNP Q9KL03
A	25	LEU	ARG	engineered mutation	UNP Q9KL03
A	26	ALA	ASN	engineered mutation	UNP Q9KL03
A	27	ARG	ILE	engineered mutation	UNP Q9KL03
A	28	TYR	MET	engineered mutation	UNP Q9KL03
A	29	GLU	SER	engineered mutation	UNP Q9KL03
A	31	MET	TRP	engineered mutation	UNP Q9KL03
A	32	GLU	PHE	engineered mutation	UNP Q9KL03
A	34	GLN	GLU	engineered mutation	UNP Q9KL03
B	23	LYS	ASN	engineered mutation	UNP Q9KL03
B	24	GLU	ASN	engineered mutation	UNP Q9KL03
B	25	LEU	ARG	engineered mutation	UNP Q9KL03
B	26	ALA	ASN	engineered mutation	UNP Q9KL03
B	27	ARG	ILE	engineered mutation	UNP Q9KL03
B	28	TYR	MET	engineered mutation	UNP Q9KL03
B	29	GLU	SER	engineered mutation	UNP Q9KL03
B	31	MET	TRP	engineered mutation	UNP Q9KL03
B	32	GLU	PHE	engineered mutation	UNP Q9KL03
B	34	GLN	GLU	engineered mutation	UNP Q9KL03
C	23	LYS	ASN	engineered mutation	UNP Q9KL03
C	24	GLU	ASN	engineered mutation	UNP Q9KL03
C	25	LEU	ARG	engineered mutation	UNP Q9KL03

*Continued on next page...*

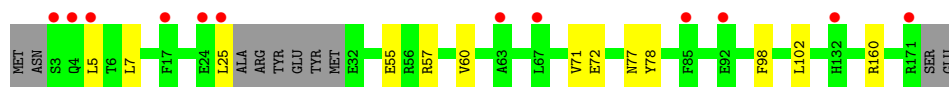
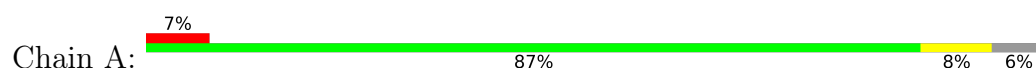
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	26	ALA	ASN	engineered mutation	UNP Q9KL03
C	27	ARG	ILE	engineered mutation	UNP Q9KL03
C	28	TYR	MET	engineered mutation	UNP Q9KL03
C	29	GLU	SER	engineered mutation	UNP Q9KL03
C	31	MET	TRP	engineered mutation	UNP Q9KL03
C	32	GLU	PHE	engineered mutation	UNP Q9KL03
C	34	GLN	GLU	engineered mutation	UNP Q9KL03

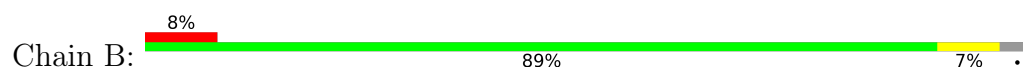
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

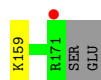
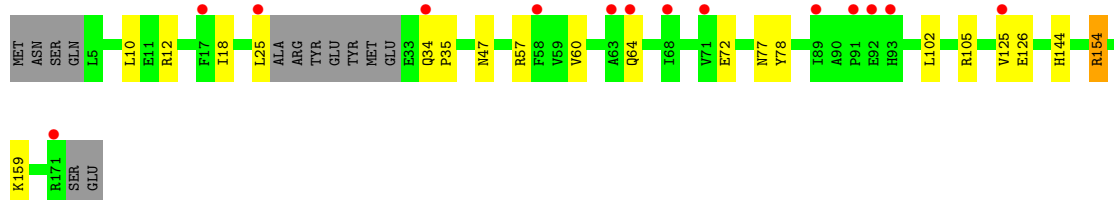
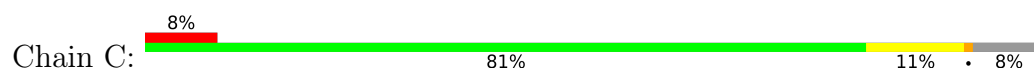
- Molecule 1: Spermidine N(1)-acetyltransferase



- Molecule 1: Spermidine N(1)-acetyltransferase



- Molecule 1: Spermidine N(1)-acetyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.90Å 134.97Å 138.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.97 – 2.58 29.97 – 2.58	Depositor EDS
% Data completeness (in resolution range)	98.0 (29.97-2.58) 98.0 (29.97-2.58)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.47 (at 2.57Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.235 , 0.278 0.235 , 0.278	Depositor DCC
$R_{free}$ test set	1156 reflections (5.33%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	54.8	Xtriage
Anisotropy	0.451	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 43.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.008 for -h,-l,-k	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8005	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.40% 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

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	0.24	0/1379	0.39	0/1859
1	B	0.24	0/1404	0.39	0/1894
1	C	0.24	0/1366	0.39	0/1842
All	All	0.24	0/4149	0.39	0/5595

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

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	1352	1308	1304	11	0
1	B	1376	1335	1329	6	0
1	C	1338	1296	1296	12	0
All	All	4066	3939	3929	28	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 4.

The worst 5 of 28 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:60:VAL:HG11	1:A:102:LEU:HD22	1.58	0.82

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:60:VAL:HG11	1:C:102:LEU:HD22	1.72	0.72
1:B:100:ARG:NH1	1:B:136:GLU:OE2	2.30	0.64
1:B:81:ARG:NH1	1:B:165:GLN:OE1	2.31	0.64
1:A:57:ARG:NH1	1:A:72:GLU:OE2	2.32	0.62

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	159/173 (92%)	154 (97%)	5 (3%)	0	100	100
1	B	162/173 (94%)	158 (98%)	4 (2%)	0	100	100
1	C	156/173 (90%)	150 (96%)	6 (4%)	0	100	100
All	All	477/519 (92%)	462 (97%)	15 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	141/156 (90%)	141 (100%)	0	100	100
1	B	144/156 (92%)	143 (99%)	1 (1%)	84	93
1	C	141/156 (90%)	139 (99%)	2 (1%)	67	84

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	426/468 (91%)	423 (99%)	3 (1%)	84	93

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	171	ARG
1	C	105	ARG
1	C	154	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides 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, 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	163/173 (94%)	0.50	12 (7%) 14 12	40, 58, 83, 92	0
1	B	166/173 (95%)	0.65	13 (7%) 13 10	39, 58, 91, 101	0
1	C	160/173 (92%)	0.72	14 (8%) 10 8	43, 60, 87, 114	0
All	All	489/519 (94%)	0.63	39 (7%) 12 10	39, 59, 90, 114	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	3	SER	5.8
1	C	63	ALA	4.2
1	A	25	LEU	4.0
1	B	68	ILE	4.0
1	C	64	GLN	3.9

### 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 monosaccharides 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.