



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 27, 2017 – 12:15 PM EST

PDB ID : 5Y9G  
Title : Crystal structure of Salmonella SafD adhesin  
Authors : Zeng, L.H.; Zhang, L.; Wang, P.R.; Meng, G.  
Deposited on : unknown  
Resolution : 2.20 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
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

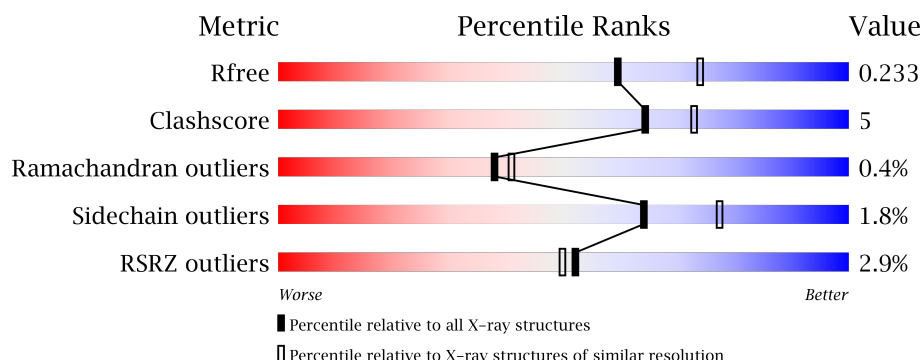
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4002 (2.20-2.20)
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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	149	<div> <div>3%</div> <div>79%</div> <div>11%</div> <div>9%</div> </div>
1	B	149	<div> <div>2%</div> <div>89%</div> <div>5%</div> <div>5%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4282 atoms, of which 2027 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 Pilin structural protein SafD, Pilus assembly protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	136	Total	C	H	N	O	S	0	0	0
			2043	653	995	189	204	2			
1	B	141	Total	C	H	N	O	S	0	0	0
			2114	676	1032	194	210	2			

There are 8 discrepancies between the modelled and reference sequences:

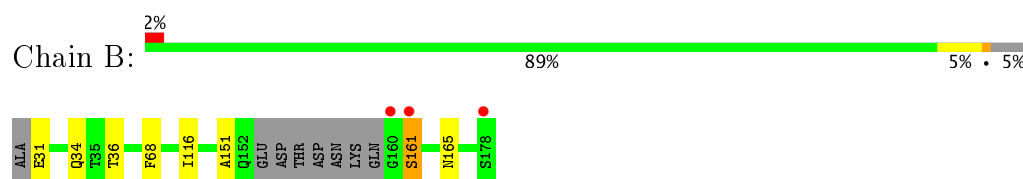
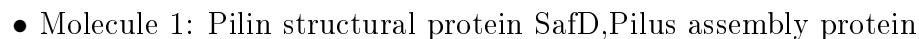
Chain	Residue	Modelled	Actual	Comment	Reference
A	156	ASP	-	linker	UNP A0A1X2WLP6
A	157	ASN	-	linker	UNP A0A1X2WLP6
A	158	LYS	-	linker	UNP A0A1X2WLP6
A	159	GLN	-	linker	UNP A0A1X2WLP6
B	156	ASP	-	linker	UNP A0A1X2WLP6
B	157	ASN	-	linker	UNP A0A1X2WLP6
B	158	LYS	-	linker	UNP A0A1X2WLP6
B	159	GLN	-	linker	UNP A0A1X2WLP6

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	62	Total	O	0	0
			62	62		
2	B	63	Total	O	0	0
			63	63		



- Molecule 1: Pilin structural protein SafD, Pilus assembly protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	32.30 Å 49.71 Å 148.83 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.15 – 2.20 47.15 – 2.20	Depositor EDS
% Data completeness (in resolution range)	90.4 (47.15-2.20) 90.4 (47.15-2.20)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.19 (at 2.20 Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.199 , 0.233 0.196 , 0.233	Depositor DCC
$R_{free}$ test set	588 reflections (5.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	19.8	Xtriage
Anisotropy	0.235	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 48.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4282	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.95% 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	0.26	0/1072	0.53	0/1456
1	B	0.27	0/1107	0.52	0/1504
All	All	0.27	0/2179	0.53	0/2960

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	1048	995	1000	15	0
1	B	1082	1032	1029	7	0
2	A	62	0	0	1	0
2	B	63	0	0	3	0
All	All	2255	2027	2029	20	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.

The worst 5 of 20 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:161:SER:O	2:B:201:HOH:O	1.82	0.96

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:171:SER:O	2:A:202:HOH:O	2.09	0.71
1:A:33:LEU:HD22	1:A:168:GLN:HG2	1.73	0.70
1:A:33:LEU:O	1:A:167:GLU:HG2	1.98	0.64
1:A:41:THR:HG23	1:A:42:TYR:CD2	2.38	0.58

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	132/149 (89%)	129 (98%)	2 (2%)	1 (1%)	22	21
1	B	137/149 (92%)	136 (99%)	1 (1%)	0	100	100
All	All	269/298 (90%)	265 (98%)	3 (1%)	1 (0%)	38	41

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	40	GLY

### 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	112/123 (91%)	110 (98%)	2 (2%)	64	77

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	115/123 (94%)	113 (98%)	2 (2%)	66	79
All	All	227/246 (92%)	223 (98%)	4 (2%)	64	77

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	84	LEU
1	A	167	GLU
1	B	31	GLU
1	B	161	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	56	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 ⓘ

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	136/149 (91%)	0.11	5 (3%) 42 40	9, 28, 52, 77	0
1	B	141/149 (94%)	-0.09	3 (2%) 64 61	8, 20, 38, 60	0
All	All	277/298 (92%)	0.01	8 (2%) 52 50	8, 24, 48, 77	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	178	SER	3.6
1	A	30	ALA	3.2
1	B	160	GLY	3.1
1	A	135	GLU	2.9
1	B	178	SER	2.7

### 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.