



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:34 PM GMT

PDB ID : 4PAY  
Title : Crystal structure of an N-terminal fragment of the Legionella pneumophila effector protein SidC.  
Authors : Horenkamp, F.A.; Mukherjee, S.; Alix, E.; Schauder, C.M.; Hubber, A.M.; Roy, C.R.; Reinisch, K.M.  
Deposited on : 2014-04-10  
Resolution : 2.77 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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

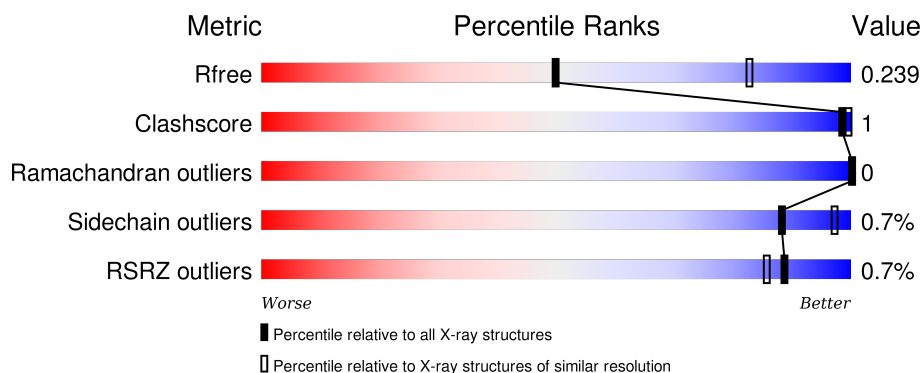
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

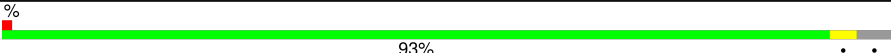
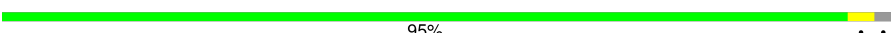
The reported resolution of this entry is 2.77 Å.

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	3004 (2.80-2.76)
Clashscore	102246	3480 (2.80-2.76)
Ramachandran outliers	100387	3423 (2.80-2.76)
Sidechain outliers	100360	3425 (2.80-2.76)
RSRZ outliers	91569	3016 (2.80-2.76)

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	610	 93% . .
1	B	610	 95% . .

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 18699 atoms, of which 9083 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 SidC, interaptin.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
1	A	588	Total	C	H	N	O	S	Se		0	2	0
			9086	2947	4440	780	913	4	2				
1	B	598	Total	C	H	N	O	S	Se		0	1	0
			9439	3033	4643	808	949	4	2				

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q5ZSK6
A	0	SER	-	expression tag	UNP Q5ZSK6
B	-1	GLY	-	expression tag	UNP Q5ZSK6
B	0	SER	-	expression tag	UNP Q5ZSK6

- Molecule 2 is BARIUM ION (three-letter code: BA) (formula: Ba).

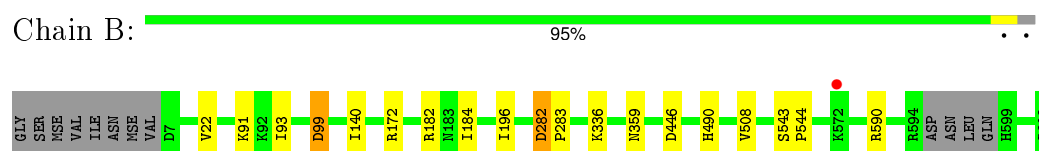
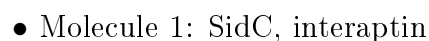
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	5	Total	Ba	0	0
			5	5		
2	A	3	Total	Ba	0	0
			3	3		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	91	Total	O	0	0
			91	91		
3	B	75	Total	O	0	0
			75	75		



- Molecule 1: SidC, interaptin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.46Å 107.58Å 177.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	92.01 – 2.77 92.01 – 2.77	Depositor EDS
% Data completeness (in resolution range)	99.8 (92.01-2.77) 95.3 (92.01-2.77)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.60 (at 2.77Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.199 , 0.246 0.194 , 0.239	Depositor DCC
$R_{free}$ test set	1911 reflections (3.84%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.5	Xtriage
Anisotropy	0.422	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 35.3	EDS
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 52258 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	18699	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: BA

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	0.42	0/4743	0.51	0/6431
1	B	0.41	0/4894	0.51	0/6619
All	All	0.42	0/9637	0.51	0/13050

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	4646	4440	4471	9	0
1	B	4796	4643	4685	9	0
2	A	3	0	0	0	0
2	B	5	0	0	0	0
3	A	91	0	0	0	0
3	B	75	0	0	0	0
All	All	9616	9083	9156	17	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 1.

The worst 5 of 17 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:91:LYS:NZ	1:B:99:ASP:OD1	2.09	0.85
1:B:182:ARG:NH2	1:B:446:ASP:OD1	2.18	0.75
1:A:182:ARG:NH2	1:A:446:ASP:OD1	2.24	0.70
1:A:343:ASP:OD2	1:B:336:LYS:NZ	2.29	0.60
1:B:22:VAL:HG12	1:B:508:VAL:HG11	1.86	0.56

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	584/610 (96%)	575 (98%)	9 (2%)	0	100	100
1	B	595/610 (98%)	584 (98%)	11 (2%)	0	100	100
All	All	1179/1220 (97%)	1159 (98%)	20 (2%)	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	502/549 (91%)	499 (99%)	3 (1%)	90	97
1	B	531/549 (97%)	527 (99%)	4 (1%)	86	96

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1033/1098 (94%)	1026 (99%)	7 (1%)	88 97

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	99	ASP
1	B	490	HIS
1	B	282	ASP
1	A	282	ASP
1	B	359	ASN

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

### 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 8 ligands modelled in this entry, 8 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 [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	586/610 (96%)	0.33	7 (1%) 81 75	22, 40, 78, 111	0
1	B	596/610 (97%)	0.31	1 (0%) 95 95	26, 43, 65, 88	0
All	All	1182/1220 (96%)	0.32	8 (0%) 89 86	22, 41, 72, 111	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	256	ILE	3.8
1	A	266[A]	GLU	3.4
1	B	572	LYS	3.3
1	A	237	TYR	2.6
1	A	279	LEU	2.4

### 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	BA	A	703	1/1	1.00	0.25	1.60	43,43,43,43	0
2	BA	B	702	1/1	1.00	0.24	0.99	53,53,53,53	0
2	BA	B	703	1/1	0.96	0.18	0.44	119,119,119,119	0
2	BA	A	702	1/1	0.92	0.14	-1.44	147,147,147,147	0
2	BA	A	701	1/1	0.94	0.14	-1.49	113,113,113,113	0
2	BA	B	704	1/1	0.99	0.24	-	72,72,72,72	0
2	BA	B	701	1/1	0.97	0.17	-	95,95,95,95	0
2	BA	B	705	1/1	0.99	0.14	-	100,100,100,100	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.