



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 06:53 pm GMT

PDB ID : 3TUL
Title : Crystal structure of N-terminal region of Type III Secretion Major Translocator SipB (residues 82-226)
Authors : Barta, M.L.; Dickenson, N.E.; Patel, M.; Keightley, J.A.; Picking, W.D.; Picking, W.L.; Geisbrecht, B.V.
Deposited on : 2011-09-16
Resolution : 2.79 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

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	:	trunk28620
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)	:	recalc28949

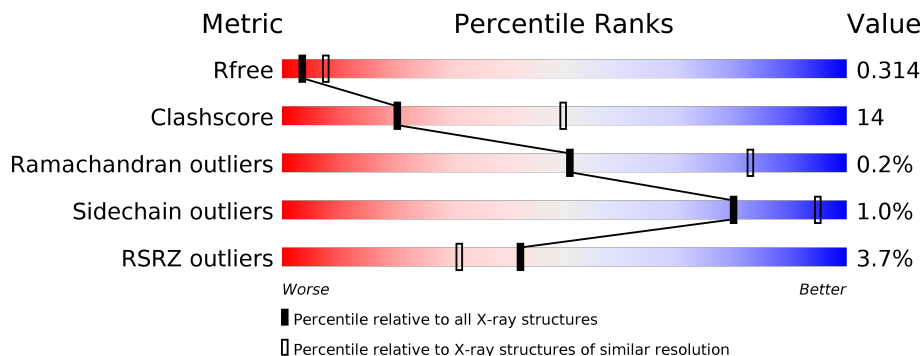
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.79 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	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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 ≥ 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	158	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>23%</div> <div>14%</div> </div> </div>
1	B	158	<div> <div>3%</div> <div> <div></div> <div>70%</div> <div>14%</div> <div>•</div> <div>16%</div> </div> </div>
1	C	158	<div> <div>5%</div> <div> <div></div> <div>56%</div> <div>25%</div> <div>•</div> <div>18%</div> </div> </div>
1	D	158	<div> <div>%</div> <div> <div></div> <div>58%</div> <div>19%</div> <div>23%</div> </div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 3919 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 Cell invasion protein sipB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	136	Total	C	N	O	Se	0	0	0
			1013	626	169	215	3			
1	B	133	Total	C	N	O	Se	0	0	0
			990	612	166	209	3			
1	C	130	Total	C	N	O	Se	0	0	0
			968	598	162	205	3			
1	D	122	Total	C	N	O	Se	0	0	0
			905	561	152	190	2			

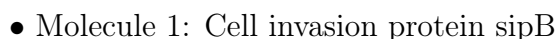
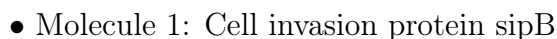
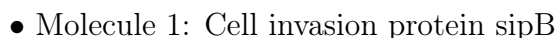
There are 4 discrepancies between the modelled and reference sequences:

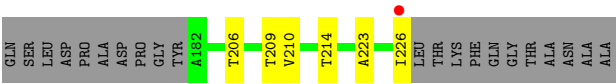
Chain	Residue	Modelled	Actual	Comment	Reference
A	80	GLY	-	EXPRESSION TAG	UNP Q56019
B	80	GLY	-	EXPRESSION TAG	UNP Q56019
C	80	GLY	-	EXPRESSION TAG	UNP Q56019
D	80	GLY	-	EXPRESSION TAG	UNP Q56019

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	7	Total	O	0	0
			7	7		
2	B	15	Total	O	0	0
			15	15		
2	C	7	Total	O	0	0
			7	7		
2	D	14	Total	O	0	0
			14	14		

- Molecule 1: Cell invasion protein sipB





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 2 21	Depositor
Cell constants a, b, c, α , β , γ	159.15Å 51.14Å 84.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	31.94 – 2.79 31.94 – 2.79	Depositor EDS
% Data completeness (in resolution range)	94.3 (31.94-2.79) 94.4 (31.94-2.79)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.42 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.298 , 0.316 0.294 , 0.314	Depositor DCC
R_{free} test set	840 reflections (4.99%)	DCC
Wilson B-factor (Å ²)	52.6	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 38.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	3919	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 36.69 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 4.8053e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.46	0/1012	0.62	0/1351
1	B	0.51	0/989	0.65	0/1320
1	C	0.45	0/967	0.60	0/1291
1	D	0.45	0/905	0.61	0/1211
All	All	0.47	0/3873	0.62	0/5173

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	1013	0	1033	29	0
1	B	990	0	1011	25	0
1	C	968	0	987	46	0
1	D	905	0	926	31	0
2	A	7	0	0	0	0
2	B	15	0	0	0	0
2	C	7	0	0	0	0
2	D	14	0	0	0	0
All	All	3919	0	3957	110	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 14.

The worst 5 of 110 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:D:149:LYS:O	1:D:152:THR:HG22	1.71	0.88
1:C:127:LYS:HG2	1:C:128:GLU:N	2.03	0.73
1:C:206:THR:O	1:C:209:THR:HG22	1.89	0.71
1:A:91:LYS:HB3	1:B:93:MSE:HE3	1.74	0.69
1:C:182:ALA:O	1:C:185:GLU:HB2	1.91	0.69

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	130/158 (82%)	120 (92%)	10 (8%)	0	100	100
1	B	127/158 (80%)	119 (94%)	8 (6%)	0	100	100
1	C	124/158 (78%)	117 (94%)	6 (5%)	1 (1%)	22	55
1	D	116/158 (73%)	108 (93%)	8 (7%)	0	100	100
All	All	497/632 (79%)	464 (93%)	32 (6%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	127	LYS

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	104/116 (90%)	102 (98%)	2 (2%)	62	89
1	B	101/116 (87%)	100 (99%)	1 (1%)	80	95
1	C	99/116 (85%)	98 (99%)	1 (1%)	80	95
1	D	91/116 (78%)	91 (100%)	0	100	100
All	All	395/464 (85%)	391 (99%)	4 (1%)	80	95

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

Mol	Chain	Res	Type
1	A	115	ILE
1	A	151	ASP
1	B	85	LEU
1	C	120	GLU

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 ⓘ

There are no ligands in this entry.

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, 95th 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 ²)	Q < 0.9
1	A	133/158 (84%)	0.22	5 (3%)	41 30	35, 66, 105, 132	0
1	B	130/158 (82%)	0.20	4 (3%)	49 38	26, 59, 101, 126	0
1	C	127/158 (80%)	0.25	8 (6%)	21 13	44, 72, 106, 138	0
1	D	120/158 (75%)	0.21	2 (1%)	70 63	33, 63, 92, 127	0
All	All	510/632 (80%)	0.22	19 (3%)	42 31	26, 65, 103, 138	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	170	LYS	5.5
1	A	123	ILE	5.0
1	C	128	GLU	4.0
1	A	120	GLU	4.0
1	C	166	GLN	4.0

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.