wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 05:01 pm BST

PDB ID	:	5I5F
Title	:	Salmonella global domain 191
Authors	:	Dong, C.; Dong, H.
Deposited on	:	2016-02-15
Resolution	:	1.84 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

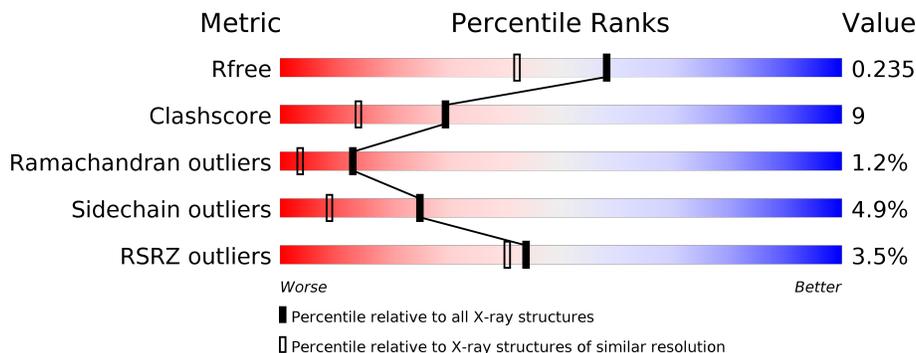|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.84 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	Percentile Ranks	Value
Rfree		0.235
Clashscore		9
Ramachandran outliers		1.2%
Sidechain outliers		4.9%
RSRZ outliers		3.5%

Worse ← → *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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 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 <=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	586	2% 46% 10% ·· 42%
1	B	586	3% 46% 9% ·· 42%
1	C	586	% 47% 9% · 42%
1	D	586	2% 48% 8% · 42%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 11684 atoms, of which 0 are hydrogens and 0 are deuteriums.
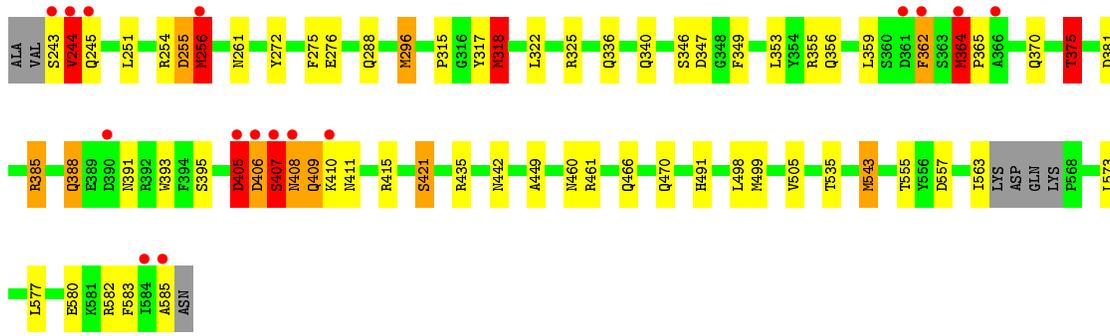
In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

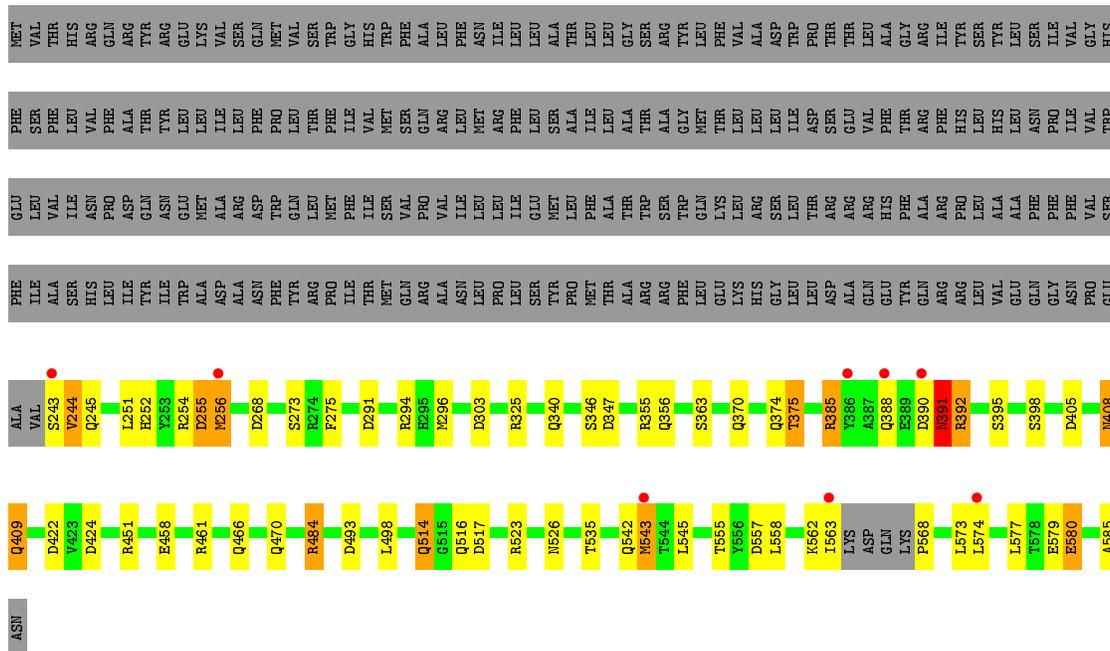- Molecule 1 is a protein called Inner membrane protein YejM.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	339	Total 2681	C 1690	N 468	O 515	S 8	0	0	0
1	B	339	Total 2681	C 1690	N 468	O 515	S 8	0	0	0
1	C	339	Total 2681	C 1690	N 468	O 515	S 8	0	0	0
1	D	339	Total 2681	C 1690	N 468	O 515	S 8	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	285	Total 285	O 285	0	0
2	B	187	Total 187	O 187	0	0
2	C	253	Total 253	O 253	0	0
2	D	235	Total 235	O 235	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: Inner membrane protein YejM

• Molecule 1: Inner membrane protein YejM

• Molecule 1: Inner membrane protein YejM

Chain C:

• Molecule 1: Inner membrane protein YejM

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.57Å 195.88Å 70.09Å 90.00° 95.75° 90.00°	Depositor
Resolution (Å)	65.70 – 1.84 65.70 – 1.84	Depositor EDS
% Data completeness (in resolution range)	86.2 (65.70-1.84) 86.2 (65.70-1.84)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.88 (at 1.84Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.165 , 0.244 0.171 , 0.235	Depositor DCC
R_{free} test set	5005 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å^2)	17.4	Xtriage
Anisotropy	0.911	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 46.4	EDS
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	11684	wwPDB-VP
Average B, all atoms (Å^2)	36.0	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 77.82 % 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 7.8717e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $<|L|>$, $<L^2>$ 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	1.28	11/2744 (0.4%)	1.09	22/3738 (0.6%)
1	B	1.71	12/2744 (0.4%)	1.57	26/3738 (0.7%)
1	C	1.34	6/2744 (0.2%)	1.17	25/3738 (0.7%)
1	D	1.19	6/2744 (0.2%)	1.24	20/3738 (0.5%)
All	All	1.40	35/10976 (0.3%)	1.28	93/14952 (0.6%)

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	B	0	4
1	C	0	1
1	D	0	2
All	All	0	7

The worst 5 of 35 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	255	ASP	C-N	-49.91	0.19	1.34
1	B	317	TYR	C-N	-42.27	0.36	1.34
1	C	255	ASP	C-N	-33.40	0.57	1.34
1	B	256	MET	C-N	-31.93	0.75	1.33
1	A	255	ASP	C-N	-31.48	0.61	1.34

The worst 5 of 93 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	256	MET	O-C-N	-42.37	51.18	123.20
1	B	317	TYR	O-C-N	-28.15	77.65	122.70
1	B	255	ASP	CA-C-N	-25.95	60.10	117.20
1	D	255	ASP	C-N-CA	-24.42	60.64	121.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	256	MET	CA-C-N	23.88	163.96	116.20

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	255	ASP	Mainchain
1	B	256	MET	Mainchain
1	B	318	MET	Mainchain
1	B	407	SER	Peptide
1	C	391	ASN	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	2681	0	2580	76	1
1	B	2681	0	2579	47	0
1	C	2681	0	2581	48	0
1	D	2681	0	2581	32	0
2	A	285	0	0	23	1
2	B	187	0	0	16	0
2	C	253	0	0	21	0
2	D	235	0	0	11	0
All	All	11684	0	10321	199	1

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 9.

The worst 5 of 199 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:C:255:ASP:CA	1:C:256:MET:N	1.72	1.48
1:C:255:ASP:C	1:C:256:MET:CA	1.76	1.48
1:A:255:ASP:C	1:A:256:MET:CA	1.86	1.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:255:ASP:CA	1:A:256:MET:N	1.81	1.42
1:C:255:ASP:O	1:C:256:MET:N	1.69	1.24

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:256:MET:SD	2:A:864:HOH:O[1_455]	2.15	0.05

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	335/586 (57%)	316 (94%)	16 (5%)	3 (1%)	17	6
1	B	335/586 (57%)	311 (93%)	18 (5%)	6 (2%)	8	1
1	C	335/586 (57%)	319 (95%)	13 (4%)	3 (1%)	17	6
1	D	335/586 (57%)	321 (96%)	10 (3%)	4 (1%)	13	3
All	All	1340/2344 (57%)	1267 (95%)	57 (4%)	16 (1%)	13	3

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	318	MET
1	B	406	ASP
1	B	409	GLN
1	C	409	GLN
1	D	409	GLN

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	291/511 (57%)	278 (96%)	13 (4%)	27	10
1	B	291/511 (57%)	276 (95%)	15 (5%)	23	8
1	C	291/511 (57%)	276 (95%)	15 (5%)	23	8
1	D	291/511 (57%)	277 (95%)	14 (5%)	25	9
All	All	1164/2044 (57%)	1107 (95%)	57 (5%)	25	9

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	573	LEU
1	C	294	ARG
1	D	535	THR
1	B	577	LEU
1	C	244	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	470	GLN
1	C	261	ASN
1	D	486	ASN
1	B	483	GLN
1	B	491	HIS

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 (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	6
1	D	4
1	A	3
1	C	2

The worst 5 of 15 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	318:MET	C	319:ASP	N	1.19
1	B	296:MET	C	297:SER	N	1.17
1	D	498:LEU	C	499:MET	N	1.15
1	A	256:MET	C	257:GLY	N	1.13
1	A	498:LEU	C	499:MET	N	1.11

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	339/586 (57%)	-0.22	12 (3%)	44	40	17, 28, 59, 91	0
1	B	339/586 (57%)	-0.03	16 (4%)	31	28	21, 36, 73, 119	0
1	C	339/586 (57%)	-0.15	8 (2%)	59	57	18, 30, 65, 113	0
1	D	339/586 (57%)	-0.19	11 (3%)	47	44	19, 31, 65, 107	0
All	All	1356/2344 (57%)	-0.15	47 (3%)	44	40	17, 31, 67, 119	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	243	SER	9.4
1	C	386	TYR	7.9
1	B	364	MET	5.7
1	A	364	MET	5.2
1	A	243	SER	4.9

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 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.