Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2020 – 01:52 am BST

PDB ID	:	5J8Y
Title	:	Crystal structure of the Scm-SAM and Sfmbt-SAM heterodimer
Authors	:	Frey, F.; Benda, C.; Mueller, J.
Deposited on	:	2016-04-08
Resolution	:	1.98 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

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

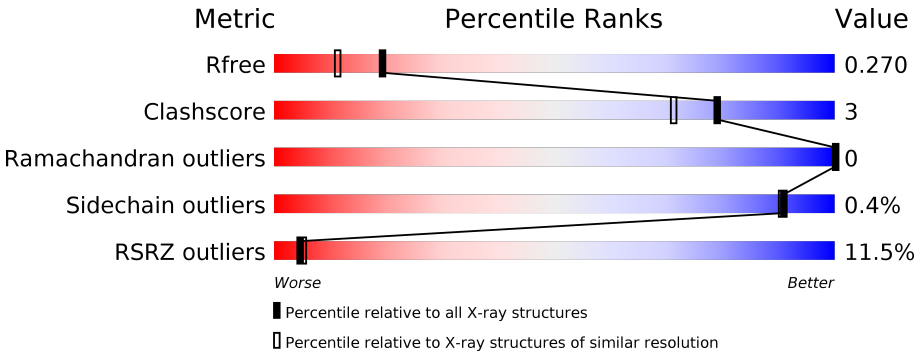
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.98 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	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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	80	5%, 94%, • •
1	B	80	8%, 91%, 5% • •
2	C	89	16%, 66%, 11%, 22%
2	D	89	11%, 76%, •, 21%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2495 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 Polycomb protein Scm.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	78	Total	C	N	O	S	0	0	0
			619	388	107	118	6			
1	B	78	Total	C	N	O	S	0	0	0
			619	388	107	118	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	798	GLY	-	expression tag	UNP Q9VHA0
A	799	PRO	-	expression tag	UNP Q9VHA0
A	800	ASP	-	expression tag	UNP Q9VHA0
A	801	SER	-	expression tag	UNP Q9VHA0
A	802	MET	-	expression tag	UNP Q9VHA0
A	855	GLU	LEU	conflict	UNP Q9VHA0
A	859	GLU	LEU	conflict	UNP Q9VHA0
B	798	GLY	-	expression tag	UNP Q9VHA0
B	799	PRO	-	expression tag	UNP Q9VHA0
B	800	ASP	-	expression tag	UNP Q9VHA0
B	801	SER	-	expression tag	UNP Q9VHA0
B	802	MET	-	expression tag	UNP Q9VHA0
B	855	GLU	LEU	conflict	UNP Q9VHA0
B	859	GLU	LEU	conflict	UNP Q9VHA0

- Molecule 2 is a protein called Polycomb protein Sfmbt.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	69	Total	C	N	O	S	4	0	0
			544	345	94	99	6			
2	D	70	Total	C	N	O	S	0	1	0
			558	352	97	103	6			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1132	GLY	-	expression tag	UNP Q9VK33
C	1133	PRO	-	expression tag	UNP Q9VK33
C	1134	ASP	-	expression tag	UNP Q9VK33
C	1135	SER	-	expression tag	UNP Q9VK33
C	1136	MET	-	expression tag	UNP Q9VK33
D	1132	GLY	-	expression tag	UNP Q9VK33
D	1133	PRO	-	expression tag	UNP Q9VK33
D	1134	ASP	-	expression tag	UNP Q9VK33
D	1135	SER	-	expression tag	UNP Q9VK33
D	1136	MET	-	expression tag	UNP Q9VK33

- Molecule 3 is water.

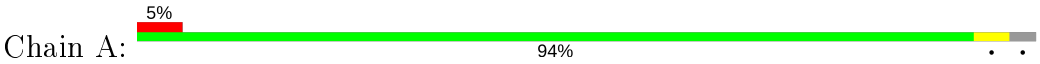
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	39	Total O 39 39	0	0
3	B	46	Total O 46 46	0	0
3	C	36	Total O 36 36	0	0
3	D	34	Total O 34 34	0	0

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.

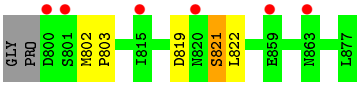
- Molecule 1: Polycomb protein Scm

Chain A:



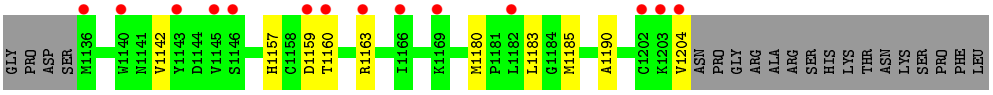
- Molecule 1: Polycomb protein Scm

Chain B:



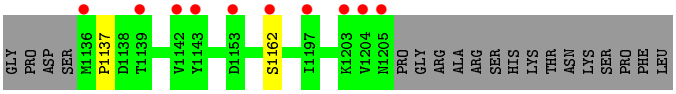
- Molecule 2: Polycomb protein Sfmbt

Chain C:



- Molecule 2: Polycomb protein Sfmbt

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	49.96Å 53.97Å 61.45Å 90.00° 109.22° 90.00°	Depositor
Resolution (Å)	47.17 – 1.98 47.17 – 1.97	Depositor EDS
% Data completeness (in resolution range)	96.1 (47.17-1.98) 96.9 (47.17-1.97)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.96 (at 1.97Å)	Xtriage
Refinement program	PHENIX 1.9_1678	Depositor
R, R_{free}	0.255 , 0.277 0.264 , 0.270	Depositor DCC
R_{free} test set	1060 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	23.8	Xtriage
Anisotropy	0.880	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	2495	wwPDB-VP
Average B, all atoms (Å^2)	29.0	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 53.84 % 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 3.9765e-05. 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

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	0.93	1/627 (0.2%)	0.59	0/842
1	B	0.87	1/627 (0.2%)	0.68	1/842 (0.1%)
2	C	1.02	0/552	0.80	0/744
2	D	0.86	3/566 (0.5%)	0.72	0/763
All	All	0.92	5/2372 (0.2%)	0.70	1/3191 (0.0%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	844	SER	CB-OG	-5.80	1.34	1.42
1	B	803	PRO	N-CD	5.42	1.55	1.47
2	D	1137	PRO	N-CD	5.30	1.55	1.47
2	D	1162[A]	SER	C-N	-5.29	1.21	1.34
2	D	1162[B]	SER	C-N	-5.29	1.21	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	802	MET	C-N-CD	5.02	138.94	128.40

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	619	0	623	1	0
1	B	619	0	623	2	0
2	C	544	0	565	10	0
2	D	558	0	575	0	0
3	A	39	0	0	0	0
3	B	46	0	0	0	0
3	C	36	0	0	1	0
3	D	34	0	0	0	0
All	All	2495	0	2386	13	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 3.

All (13) 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:859:GLU:HG3	1:A:877:LEU:HD13	1.84	0.59
1:B:819:ASP:HB3	1:B:822:LEU:HD12	1.89	0.54
2:C:1142:VAL:HG23	3:C:1328:HOH:O	2.10	0.51
2:C:1159:ASP:HB3	2:C:1163:ARG:HH12	1.78	0.48
1:B:819:ASP:OD2	1:B:821:SER:HB3	2.13	0.48
2:C:1157:HIS:ND1	2:C:1183:LEU:HD22	2.31	0.46
2:C:1160:THR:HA	2:C:1163:ARG:NH2	2.30	0.46
2:C:1183:LEU:HD12	2:C:1190:ALA:HA	1.99	0.44
2:C:1180:MET:HE3	2:C:1185:MET:HA	2.01	0.42
2:C:1160:THR:N	2:C:1163:ARG:NH2	2.68	0.42
2:C:1204:VAL:CG1	2:C:1204:VAL:O	2.69	0.41
2:C:1160:THR:CA	2:C:1163:ARG:NH2	2.84	0.41
2:C:1204:VAL:O	2:C:1204:VAL:HG12	2.21	0.40

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	76/80 (95%)	75 (99%)	1 (1%)	0	100	100
1	B	76/80 (95%)	75 (99%)	1 (1%)	0	100	100
2	C	67/89 (75%)	65 (97%)	2 (3%)	0	100	100
2	D	69/89 (78%)	68 (99%)	1 (1%)	0	100	100
All	All	288/338 (85%)	283 (98%)	5 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains

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	69/70 (99%)	69 (100%)	0	100	100
1	B	69/70 (99%)	68 (99%)	1 (1%)	67	62
2	C	63/80 (79%)	63 (100%)	0	100	100
2	D	65/80 (81%)	65 (100%)	0	100	100
All	All	266/300 (89%)	265 (100%)	1 (0%)	91	90

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	821	SER

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

6.1 Protein, DNA and RNA chains

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^{th} 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	78/80 (97%)	0.86	4 (5%) 28 30	18, 27, 40, 72	0
1	B	78/80 (97%)	0.87	6 (7%) 13 15	18, 26, 40, 73	0
2	C	69/89 (77%)	1.32	14 (20%) 1 0	21, 32, 41, 50	1 (1%)
2	D	70/89 (78%)	1.24	10 (14%) 2 2	21, 28, 41, 77	0
All	All	295/338 (87%)	1.06	34 (11%) 4 5	18, 28, 41, 77	1 (0%)

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	1205	ASN	8.1
1	A	800	ASP	7.3
1	B	800	ASP	6.5
2	C	1202	CYS	4.7
2	C	1146	SER	4.2
2	C	1203	LYS	4.1
2	C	1166	ILE	4.0
2	C	1159	ASP	3.3
1	B	820	ASN	3.2
2	D	1136	MET	3.1
2	D	1162[A]	SER	3.0
1	A	802	MET	2.9
2	C	1136	MET	2.7
2	C	1145	VAL	2.6
1	A	812	ILE	2.4
1	A	820	ASN	2.4
2	C	1204	VAL	2.4
2	D	1142	VAL	2.4
2	C	1140	TRP	2.3
1	B	801	SER	2.3
2	D	1197	ILE	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	D	1139	THR	2.3
2	C	1182	LEU	2.3
1	B	815	ILE	2.2
2	D	1153	ASP	2.2
2	C	1143	TYR	2.2
2	C	1169	LYS	2.2
2	D	1203	LYS	2.2
2	D	1204	VAL	2.2
2	D	1143	TYR	2.1
1	B	863	ASN	2.1
2	C	1160	THR	2.1
1	B	859	GLU	2.0
2	C	1163	ARG	2.0

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.