wwPDB X-ray Structure Validation Summary Report (i)

Apr 12, 2021 – 10:21 AM EDT

PDB ID	:	7M1A
Title	:	SusE-like protein BT2857
Authors	:	Suits, M.D.L.
Deposited on	:	2021-03-12
Resolution	:	1.42 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:

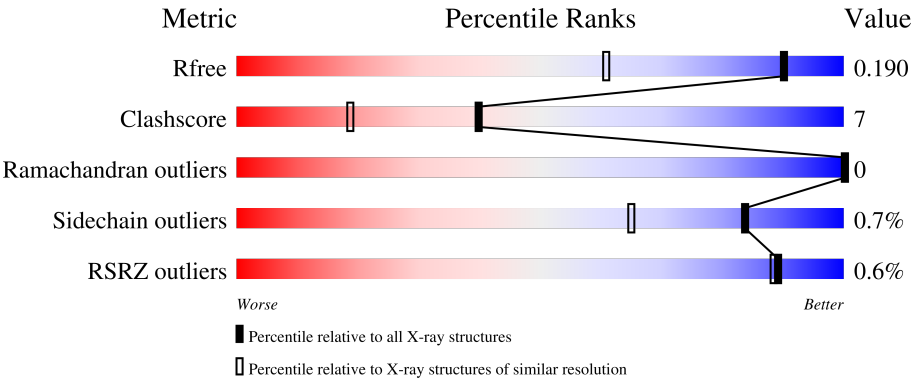
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.18
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.18

1 Overall quality at a glance (i)

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

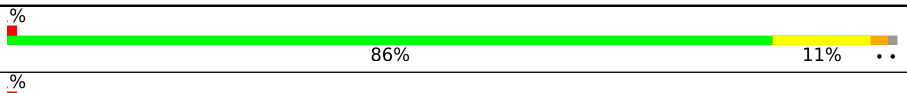
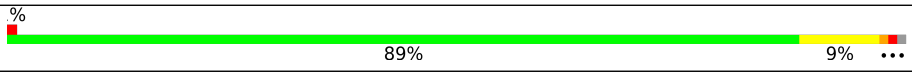
The reported resolution of this entry is 1.42 Å.

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	2579 (1.44-1.40)
Clashscore	141614	2696 (1.44-1.40)
Ramachandran outliers	138981	2632 (1.44-1.40)
Sidechain outliers	138945	2631 (1.44-1.40)
RSRZ outliers	127900	2528 (1.44-1.40)

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 of 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	AAA	179	
1	BBB	179	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	EDO	BBB	404	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3283 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.

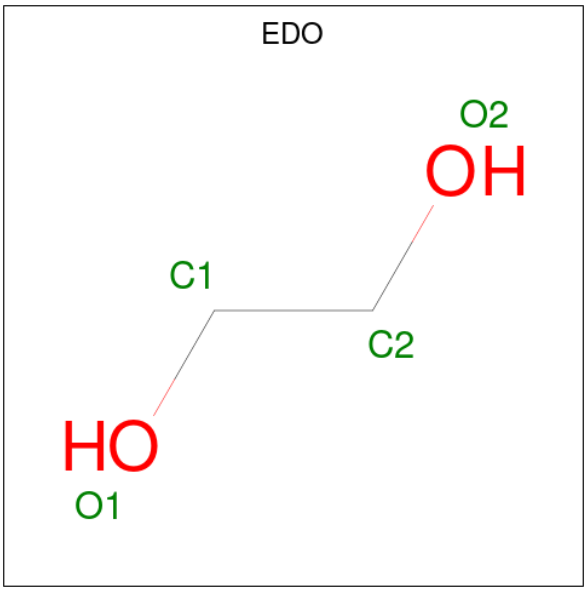
- Molecule 1 is a protein called Galactose-binding like protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	177	Total	C	N	O	S	Se	0	0	0
			1423	902	241	271	3	6			
1	BBB	177	Total	C	N	O	S	Se	0	0	0
			1423	902	241	271	3	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	220	MSE	-	initiating methionine	UNP Q8A3U7
BBB	220	MSE	-	initiating methionine	UNP Q8A3U7

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	1	Total C O 4 2 2	0	0
2	AAA	1	Total C O 4 2 2	0	0
2	BBB	1	Total C O 4 2 2	0	0
2	BBB	1	Total C O 4 2 2	0	0
2	BBB	1	Total C O 4 2 2	0	0
2	BBB	1	Total C O 4 2 2	0	0
2	BBB	1	Total C O 4 2 2	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total Ca 1 1	0	0
3	BBB	1	Total Ca 1 1	0	0

- Molecule 4 is water.

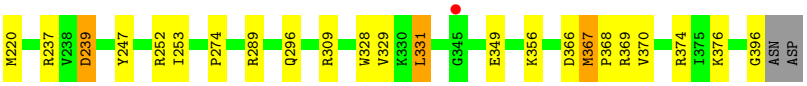
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	207	Total O 207 207	0	0
4	BBB	196	Total O 196 196	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

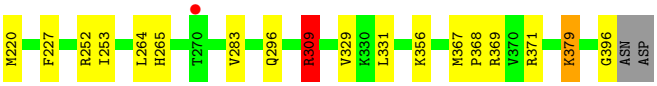
• Molecule 1: Galactose-binding like protein

Chain AAA:



• Molecule 1: Galactose-binding like protein

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	59.30Å 61.08Å 61.76Å 90.00° 101.59° 90.00°	Depositor
Resolution (Å)	43.02 – 1.42 42.98 – 1.42	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.02-1.42) 99.5 (42.98-1.42)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.23 (at 1.42Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.167 , 0.199 0.163 , 0.190	Depositor DCC
R_{free} test set	3810 reflections (4.70%)	wwPDB-VP
Wilson B-factor (Å^2)	9.7	Xtriage
Anisotropy	0.246	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 45.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.007 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3283	wwPDB-VP
Average B, all atoms (Å^2)	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.23% of the height of the origin peak. No significant pseudotranslation is detected.*

[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

Bond lengths and bond angles in the following residue types are not validated in this section: CA, EDO

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	AAA	0.89	1/1455 (0.1%)	1.20	12/1962 (0.6%)
1	BBB	0.83	0/1455	1.14	5/1962 (0.3%)
All	All	0.86	1/2910 (0.0%)	1.17	17/3924 (0.4%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	AAA	349	GLU	CD-OE2	10.45	1.37	1.25

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	367	MSE	CG-SE-CE	10.30	121.55	98.90
1	AAA	309	ARG	NE-CZ-NH1	9.55	125.08	120.30
1	AAA	309	ARG	NE-CZ-NH2	-7.92	116.34	120.30
1	AAA	289	ARG	NE-CZ-NH2	7.08	123.84	120.30
1	BBB	396	GLY	CA-C-O	-7.02	107.97	120.60

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	AAA	1423	0	1360	19	2
1	BBB	1423	0	1359	21	2
2	AAA	12	0	18	0	0
2	BBB	20	0	30	7	0
3	AAA	1	0	0	0	0
3	BBB	1	0	0	0	0
4	AAA	207	0	0	5	2
4	BBB	196	0	0	1	2
All	All	3283	0	2767	41	4

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 7.

The worst 5 of 41 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:AAA:247:TYR:HD1	4:AAA:654:HOH:O	1.31	1.10
1:BBB:331:LEU:HD23	1:BBB:367:MSE:CE	1.87	1.04
1:BBB:331:LEU:CD2	1:BBB:367:MSE:HE2	1.89	1.03
1:AAA:247:TYR:CD1	4:AAA:654:HOH:O	2.09	1.00
1:BBB:331:LEU:HD23	1:BBB:367:MSE:HE2	0.94	0.92

All (4) 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:AAA:220:MSE:N	4:BBB:564:HOH:O[2_555]	0.33	1.87
1:BBB:220:MSE:N	4:AAA:520:HOH:O[2_645]	0.41	1.79
1:AAA:220:MSE:CA	4:BBB:564:HOH:O[2_555]	1.56	0.64
1:BBB:220:MSE:CA	4:AAA:520:HOH:O[2_645]	1.65	0.55

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	AAA	175/179 (98%)	173 (99%)	2 (1%)	0	100	100
1	BBB	175/179 (98%)	173 (99%)	2 (1%)	0	100	100
All	All	350/358 (98%)	346 (99%)	4 (1%)	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	AAA	153/149 (103%)	153 (100%)	0	100	100
1	BBB	153/149 (103%)	151 (99%)	2 (1%)	69	41
All	All	306/298 (103%)	304 (99%)	2 (1%)	84	65

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	309	ARG
1	BBB	379	LYS

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

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	EDO	AAA	401	-	3,3,3	0.42	0	2,2,2	1.26	0
2	EDO	BBB	401	-	3,3,3	0.61	0	2,2,2	0.94	0
2	EDO	BBB	404	-	3,3,3	0.42	0	2,2,2	0.36	0
2	EDO	BBB	403	-	3,3,3	0.47	0	2,2,2	1.24	0
2	EDO	BBB	405	-	3,3,3	0.28	0	2,2,2	0.29	0
2	EDO	BBB	402	-	3,3,3	0.25	0	2,2,2	2.52	1 (50%)
2	EDO	AAA	403	-	3,3,3	0.56	0	2,2,2	1.01	0
2	EDO	AAA	402	-	3,3,3	0.28	0	2,2,2	0.89	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	AAA	401	-	-	0/1/1/1	-
2	EDO	BBB	401	-	-	0/1/1/1	-
2	EDO	BBB	404	-	-	1/1/1/1	-
2	EDO	BBB	403	-	-	1/1/1/1	-
2	EDO	BBB	405	-	-	1/1/1/1	-
2	EDO	BBB	402	-	-	0/1/1/1	-
2	EDO	AAA	403	-	-	1/1/1/1	-
2	EDO	AAA	402	-	-	0/1/1/1	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	BBB	402	EDO	O2-C2-C1	-3.56	86.32	111.91

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	BBB	404	EDO	O1-C1-C2-O2
2	BBB	403	EDO	O1-C1-C2-O2
2	AAA	403	EDO	O1-C1-C2-O2
2	BBB	405	EDO	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	BBB	404	EDO	5	0
2	BBB	405	EDO	1	0
2	BBB	402	EDO	2	0

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

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	AAA	171/179 (95%)	0.37	1 (0%)	89	88	6, 11, 20, 44	0
1	BBB	171/179 (95%)	0.32	1 (0%)	89	88	6, 11, 19, 32	0
All	All	342/358 (95%)	0.35	2 (0%)	89	88	6, 11, 20, 44	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	270	THR	2.6
1	AAA	345	GLY	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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. The B-factors column lists the minimum, median, 95^{th} 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	B-factors(Å^2)	Q<0.9
2	EDO	BBB	402	4/4	0.84	0.21	24,25,33,36	0
2	EDO	AAA	403	4/4	0.88	0.17	22,23,27,35	0
2	EDO	BBB	403	4/4	0.88	0.33	29,30,33,40	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	BBB	405	4/4	0.90	0.16	26,27,34,34	0
2	EDO	BBB	404	4/4	0.92	0.12	26,27,28,39	0
2	EDO	AAA	402	4/4	0.93	0.10	17,17,27,28	0
2	EDO	AAA	401	4/4	0.95	0.12	19,21,22,24	0
2	EDO	BBB	401	4/4	0.95	0.08	16,18,23,30	0
3	CA	AAA	404	1/1	1.00	0.08	8,8,8,8	0
3	CA	BBB	406	1/1	1.00	0.08	8,8,8,8	0

6.5 Other polymers (i)

There are no such residues in this entry.