wwPDB X-ray Structure Validation Summary Report (i)

Oct 25, 2022 – 08:19 pm BST

PDB ID	:	8ARM
Title	:	Crystal structure of LSSmScarlet2
Authors	:	Samygina, V.R.; Subach, O.M.; Vlaskina, A.V.; Subach, F.V.
Deposited on	:	2022-08-17
Resolution	:	1.41 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

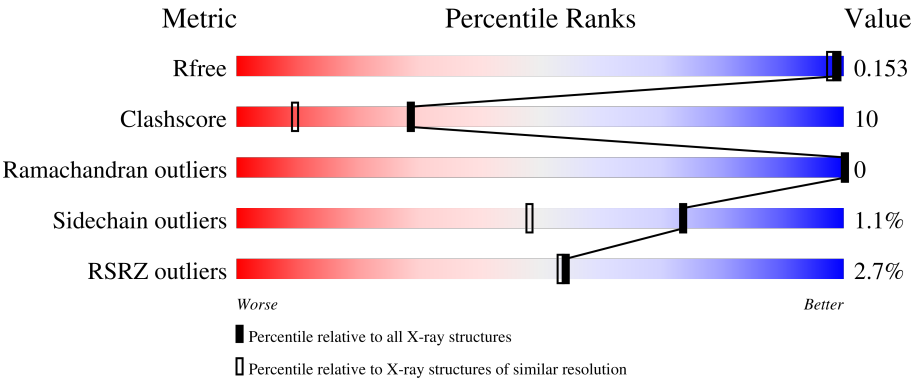
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 1.41 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	A	232	3%, 84%, 9%, • 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2274 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 LSSmScarlet2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	220	1888	1201	325	351	11	0	21	0

There are 69 discrepancies between the modelled and reference sequences:

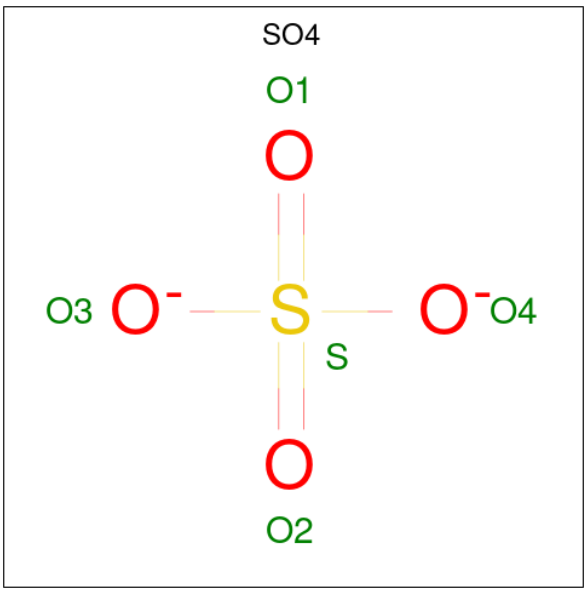
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ARG	-	expression tag	UNP Q9U6Y8
A	0	SER	-	expression tag	UNP Q9U6Y8
A	1	MET	-	expression tag	UNP Q9U6Y8
A	2	VAL	-	expression tag	UNP Q9U6Y8
A	3	SER	-	expression tag	UNP Q9U6Y8
A	4	LYS	-	expression tag	UNP Q9U6Y8
A	5	GLY	-	expression tag	UNP Q9U6Y8
A	6	GLU	-	expression tag	UNP Q9U6Y8
A	7	ALA	-	expression tag	UNP Q9U6Y8
A	18	HIS	ARG	engineered mutation	UNP Q9U6Y8
A	22	SER	THR	engineered mutation	UNP Q9U6Y8
A	23	MET	VAL	engineered mutation	UNP Q9U6Y8
A	42	THR	HIS	engineered mutation	UNP Q9U6Y8
A	43	GLN	ASN	engineered mutation	UNP Q9U6Y8
A	45	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	58	SER	ALA	engineered mutation	UNP Q9U6Y8
A	67	NRQ	GLN	chromophore	UNP Q9U6Y8
A	67	NRQ	TYR	chromophore	UNP Q9U6Y8
A	67	NRQ	GLY	chromophore	UNP Q9U6Y8
A	72	ARG	LYS	engineered mutation	UNP Q9U6Y8
A	73	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	74	PHE	TYR	engineered mutation	UNP Q9U6Y8
A	75	THR	VAL	engineered mutation	UNP Q9U6Y8
A	85	HIS	LYS	engineered mutation	UNP Q9U6Y8
A	87	GLN	LEU	engineered mutation	UNP Q9U6Y8
A	106	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	113	THR	SER	engineered mutation	UNP Q9U6Y8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	116	GLU	GLN	engineered mutation	UNP Q9U6Y8
A	119	THR	CYS	engineered mutation	UNP Q9U6Y8
A	120	LEU	PHE	engineered mutation	UNP Q9U6Y8
A	123	GLU	LYS	engineered mutation	UNP Q9U6Y8
A	126	LEU	PHE	engineered mutation	UNP Q9U6Y8
A	127	ARG	ILE	engineered mutation	UNP Q9U6Y8
A	129	THR	VAL	engineered mutation	UNP Q9U6Y8
A	133	PRO	SER	engineered mutation	UNP Q9U6Y8
A	145	LEU	TRP	engineered mutation	UNP Q9U6Y8
A	148	ASP	SER	engineered mutation	UNP Q9U6Y8
A	155	GLU	ARG	engineered mutation	UNP Q9U6Y8
A	162	ASP	GLU	engineered mutation	UNP Q9U6Y8
A	164	LYS	HIS	engineered mutation	UNP Q9U6Y8
A	165	MET	LYS	engineered mutation	UNP Q9U6Y8
A	168	ARG	LYS	engineered mutation	UNP Q9U6Y8
A	171	GLY	ASP	engineered mutation	UNP Q9U6Y8
A	174	ARG	HIS	engineered mutation	UNP Q9U6Y8
A	177	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	178	HIS	GLU	engineered mutation	UNP Q9U6Y8
A	179	VAL	PHE	engineered mutation	UNP Q9U6Y8
A	180	ARG	LYS	engineered mutation	UNP Q9U6Y8
A	181	THR	SER	engineered mutation	UNP Q9U6Y8
A	182	THR	ILE	engineered mutation	UNP Q9U6Y8
A	184	LYS	MET	engineered mutation	UNP Q9U6Y8
A	190	LEU	GLN	engineered mutation	UNP Q9U6Y8
A	191	MET	LEU	engineered mutation	UNP Q9U6Y8
A	194	ALA	TYR	engineered mutation	UNP Q9U6Y8
A	196	ASN	TYR	engineered mutation	UNP Q9U6Y8
A	199	ARG	SER	engineered mutation	UNP Q9U6Y8
A	212	VAL	ILE	engineered mutation	UNP Q9U6Y8
A	216	PHE	TYR	engineered mutation	UNP Q9U6Y8
A	219	SER	THR	engineered mutation	UNP Q9U6Y8
A	224	SER	-	expression tag	UNP Q9U6Y8
A	225	THR	-	expression tag	UNP Q9U6Y8
A	226	GLY	-	expression tag	UNP Q9U6Y8
A	227	ASP	-	expression tag	UNP Q9U6Y8
A	228	MET	-	expression tag	UNP Q9U6Y8
A	229	ASP	-	expression tag	UNP Q9U6Y8
A	230	GLU	-	expression tag	UNP Q9U6Y8
A	231	LEU	-	expression tag	UNP Q9U6Y8
A	232	TYR	-	expression tag	UNP Q9U6Y8
A	233	LYS	-	expression tag	UNP Q9U6Y8

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

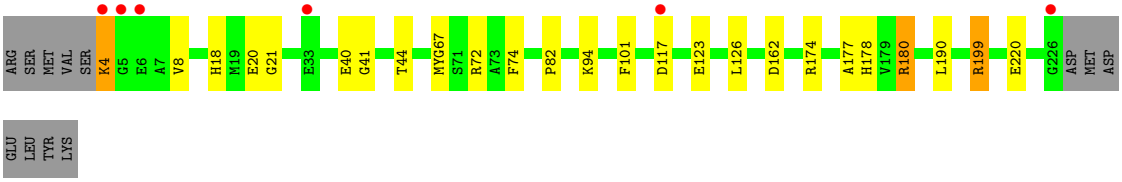
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	381	Total O 381 381	0	0

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: LSSmScarlet2

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	84.62Å 45.37Å 58.74Å 90.00° 102.09° 90.00°	Depositor
Resolution (Å)	15.00 – 1.41 14.96 – 1.41	Depositor EDS
% Data completeness (in resolution range)	97.4 (15.00-1.41) 97.5 (14.96-1.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.71 (at 1.41Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.113 , 0.152 0.113 , 0.153	Depositor DCC
R_{free} test set	2090 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	8.9	Xtriage
Anisotropy	0.296	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	2274	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.93% 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: SO4, NRQ

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	0.62	0/1963	0.81	2/2633 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	199	ARG	NE-CZ-NH2	-7.34	116.63	120.30
1	A	199	ARG	CG-CD-NE	-5.71	99.80	111.80

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	1888	0	1903	37	0
2	A	5	0	0	0	0
3	A	381	0	0	16	0
All	All	2274	0	1903	37	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 10.

The worst 5 of 37 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:94[B]:LYS:HZ2	1:A:94[B]:LYS:CB	1.09	1.51
1:A:94[B]:LYS:NZ	1:A:94[B]:LYS:HB3	1.17	1.15
1:A:123[B]:GLU:OE2	3:A:401:HOH:O	1.83	0.93
1:A:40[B]:GLU:OE1	3:A:402:HOH:O	1.90	0.87
1:A:123[B]:GLU:CG	3:A:401:HOH:O	2.21	0.85

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	236/232 (102%)	233 (99%)	3 (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	A	206/197 (105%)	203 (98%)	3 (2%)	65 36

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	4	LYS
1	A	180[A]	ARG
1	A	180[B]	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	100	ASN
1	A	178	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)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	$\#\|Z\| > 2$
1	NRQ	A	67	1	23,24,25	2.29	4 (17%)	23,32,34	2.80	3 (13%)

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
1	NRQ	A	67	1	-	2/9/31/32	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	67	NRQ	CB2-CA2	7.14	1.41	1.35
1	A	67	NRQ	CA2-C2	-6.80	1.41	1.48
1	A	67	NRQ	C2-N3	-3.15	1.32	1.39
1	A	67	NRQ	O2-C2	2.23	1.27	1.23

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	67	NRQ	CA2-C2-N3	11.71	108.91	103.37
1	A	67	NRQ	O2-C2-CA2	-4.76	128.29	130.96
1	A	67	NRQ	O3-C3-CA3	-2.95	117.47	126.39

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	67	NRQ	C2-CA2-CB2-CG2
1	A	67	NRQ	CB1-CG1-SD-CE

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand is modelled in this entry.

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	SO4	A	301	-	4,4,4	0.37	0	6,6,6	0.23	0

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

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	219/232 (94%)	0.10	6 (2%) 54 53	6, 9, 19, 53	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	226	GLY	9.2
1	A	4	LYS	5.3
1	A	5	GLY	3.3
1	A	117	ASP	2.2
1	A	6	GLU	2.2

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

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
1	NRQ	A	67	23/24	0.97	0.08	7,8,11,15	0

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	SO4	A	301	5/5	0.87	0.15	14,17,26,27	5

6.5 Other polymers (i)

There are no such residues in this entry.