Full wwPDB X-ray Structure Validation Report (i)

May 16, 2020 – 12:34 pm BST

PDB ID	:	5YSJ
Title	:	SdeA mART-C domain WT apo
Authors	:	Kim, L.; Kwon, D.H.; Song, H.K.
Deposited on	:	2017-11-14
Resolution	:	2.06 Å(reported)

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

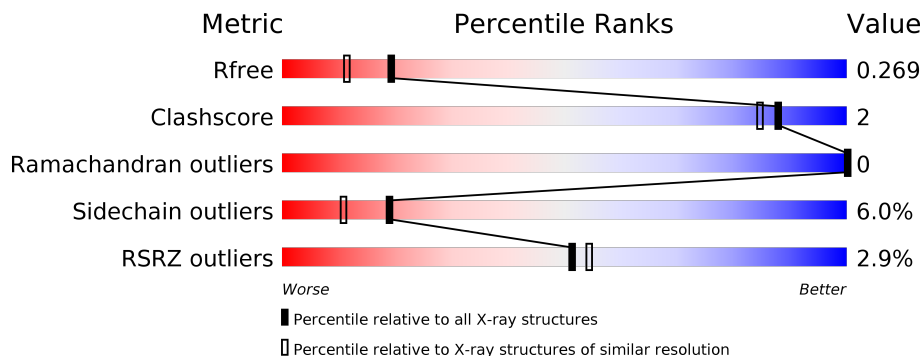|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.06 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		0.269
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		6.0%
RSRZ outliers		2.9%

■ 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	2684 (2.08-2.04)
Clashscore	141614	2801 (2.08-2.04)
Ramachandran outliers	138981	2768 (2.08-2.04)
Sidechain outliers	138945	2768 (2.08-2.04)
RSRZ outliers	127900	2646 (2.08-2.04)

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	152	2% ▮ 84% 6% · 9%
1	B	152	3% ▮ 82% 9% · 8%
1	C	152	4% ▮ 84% 7% 9%
1	D	152	% ▮ 88% 7% · 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8673 atoms, of which 4197 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 Ubiquitinating/deubiquitinating enzyme SdeA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	138	Total	C	H	N	O	S	0	0	0
			2133	690	1046	181	213	3			
1	B	140	Total	C	H	N	O	S	0	0	0
			2156	703	1050	183	217	3			
1	C	138	Total	C	H	N	O	S	0	0	0
			2142	695	1051	181	212	3			
1	D	144	Total	C	H	N	O	S	0	0	0
			2183	715	1050	192	223	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	754	GLY	-	expression tag	UNP Q5ZTK4
A	755	SER	-	expression tag	UNP Q5ZTK4
B	754	GLY	-	expression tag	UNP Q5ZTK4
B	755	SER	-	expression tag	UNP Q5ZTK4
C	754	GLY	-	expression tag	UNP Q5ZTK4
C	755	SER	-	expression tag	UNP Q5ZTK4
D	754	GLY	-	expression tag	UNP Q5ZTK4
D	755	SER	-	expression tag	UNP Q5ZTK4

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	16	Total	O	0	0
			16	16		
2	B	13	Total	O	0	0
			13	13		
2	C	8	Total	O	0	0
			8	8		
2	D	22	Total	O	0	0
			22	22		

3 Residue-property plots ⓘ

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.
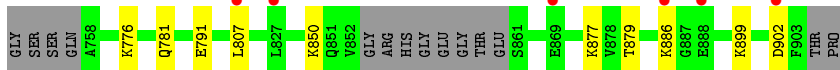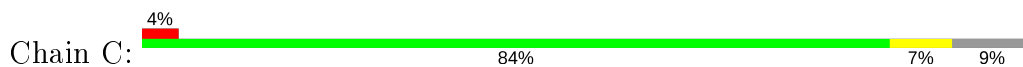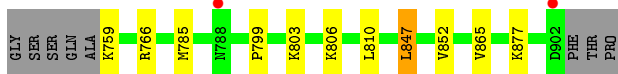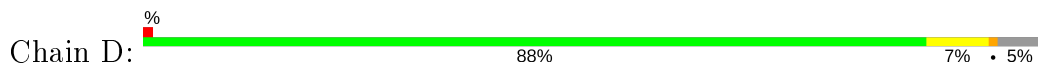
- Molecule 1: Ubiquitinating/deubiquitinating enzyme SdeA
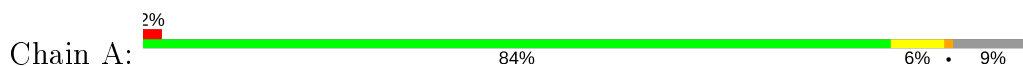
- Molecule 1: Ubiquitinating/deubiquitinating enzyme SdeA
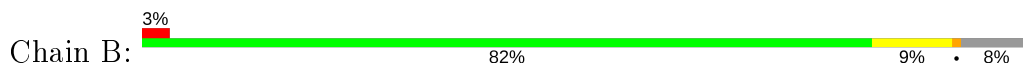
- Molecule 1: Ubiquitinating/deubiquitinating enzyme SdeA

- Molecule 1: Ubiquitinating/deubiquitinating enzyme SdeA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	43.96Å 86.68Å 71.03Å 90.00° 97.90° 90.00°	Depositor
Resolution (Å)	43.54 – 2.06 43.54 – 2.06	Depositor EDS
% Data completeness (in resolution range)	99.0 (43.54-2.06) 99.0 (43.54-2.06)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.40 (at 2.06Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.225 , 0.269 0.225 , 0.269	Depositor DCC
R_{free} test set	1996 reflections (6.16%)	wwPDB-VP
Wilson B-factor (Å^2)	25.7	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.47 , 56.6	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8673	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 41.21 % 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 2.4411e-04. 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	0.26	0/1106	0.49	0/1495
1	B	0.26	0/1126	0.50	0/1522
1	C	0.27	0/1111	0.48	0/1502
1	D	0.25	0/1154	0.53	1/1560 (0.1%)
All	All	0.26	0/4497	0.50	1/6079 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	847	LEU	C-N-CA	6.00	136.69	121.70

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	1087	1046	1093	7	0
1	B	1106	1050	1106	5	0
1	C	1091	1051	1095	2	0
1	D	1133	1050	1130	6	0
2	A	16	0	0	0	0
2	B	13	0	0	0	0
2	C	8	0	0	0	0
2	D	22	0	0	0	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4476	4197	4424	18	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 2.

All (18) 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:B:766:ARG:NH2	1:B:811:SER:O	2.28	0.66
1:D:785:MET:HE1	1:D:810:LEU:HG	1.82	0.60
1:B:766:ARG:NH2	1:B:813:MET:HG2	2.24	0.53
1:D:847:LEU:HB3	1:D:865:VAL:HG13	1.91	0.52
1:D:785:MET:HE1	1:D:810:LEU:CG	2.41	0.51
1:A:847:LEU:HB3	1:A:865:VAL:HG13	1.93	0.50
1:A:759:LYS:HA	1:A:759:LYS:HE2	1.93	0.49
1:A:813:MET:HE3	1:A:840:MET:SD	2.53	0.48
1:D:785:MET:HE1	1:D:810:LEU:CD1	2.43	0.48
1:D:799:PRO:O	1:D:803:LYS:HG2	2.13	0.48
1:B:759:LYS:HD3	1:B:759:LYS:N	2.29	0.47
1:A:860:GLU:HG3	1:A:860:GLU:H	1.41	0.44
1:A:877:LYS:HG2	1:C:879:THR:HG21	2.00	0.43
1:A:877:LYS:CG	1:C:879:THR:HG21	2.49	0.42
1:A:813:MET:CE	1:A:840:MET:SD	3.09	0.41
1:B:766:ARG:HG3	1:B:821:THR:HG22	2.02	0.41
1:B:859:THR:O	1:B:859:THR:HG23	2.21	0.40
1:D:759:LYS:HE2	1:D:759:LYS:HA	2.04	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	134/152 (88%)	129 (96%)	5 (4%)	0	100	100
1	B	136/152 (90%)	131 (96%)	5 (4%)	0	100	100
1	C	134/152 (88%)	129 (96%)	5 (4%)	0	100	100
1	D	142/152 (93%)	137 (96%)	5 (4%)	0	100	100
All	All	546/608 (90%)	526 (96%)	20 (4%)	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	124/134 (92%)	118 (95%)	6 (5%)	25	18
1	B	126/134 (94%)	115 (91%)	11 (9%)	10	4
1	C	124/134 (92%)	115 (93%)	9 (7%)	14	6
1	D	128/134 (96%)	124 (97%)	4 (3%)	40	34
All	All	502/536 (94%)	472 (94%)	30 (6%)	19	11

All (30) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	762	THR
1	A	781	GLN
1	A	804	GLN
1	A	850	LYS
1	A	860	GLU
1	A	877	LYS
1	B	759	LYS
1	B	776	LYS
1	B	789	THR
1	B	791	GLU
1	B	827	LEU
1	B	848	HIS
1	B	859	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	860	GLU
1	B	869	GLU
1	B	877	LYS
1	B	902	ASP
1	C	776	LYS
1	C	781	GLN
1	C	791	GLU
1	C	807	LEU
1	C	850	LYS
1	C	877	LYS
1	C	886	LYS
1	C	899	LYS
1	C	902	ASP
1	D	766	ARG
1	D	806	LYS
1	D	852	VAL
1	D	877	LYS

Some 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 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)

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(Å²)	Q<0.9
1	A	138/152 (90%)	0.29	3 (2%)	62	64	20, 32, 52, 74	0
1	B	140/152 (92%)	0.43	5 (3%)	42	45	16, 29, 54, 75	0
1	C	138/152 (90%)	0.40	6 (4%)	35	36	20, 33, 58, 69	0
1	D	144/152 (94%)	0.23	2 (1%)	75	76	16, 28, 49, 78	0
All	All	560/608 (92%)	0.34	16 (2%)	51	54	16, 31, 52, 78	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	853	GLY	5.0
1	B	859	THR	4.5
1	B	869	GLU	4.1
1	A	853	GLY	4.0
1	C	902	ASP	3.6
1	C	807	LEU	3.0
1	D	788	ASN	2.9
1	D	902	ASP	2.8
1	C	869	GLU	2.7
1	C	827	LEU	2.6
1	B	772	GLU	2.3
1	A	807	LEU	2.2
1	C	886	LYS	2.2
1	B	903	PHE	2.1
1	C	888	GLU	2.1
1	A	827	LEU	2.1

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.