wwPDB X-ray Structure Validation Summary Report (i)

Nov 13, 2023 – 06:10 pm GMT

PDB ID	:	7PQA
Title	:	Crystal Structure of the Ring Nuclease 0811 mutant-S12G/K169G from Sulfolobus islandicus (Sis0811)
Authors	:	Molina, R.; Jensen, A.L.G.; Marchena-Hurtado, J.; Lopez-Mendez, B.; Stella, S.; Montoya, G.
Deposited on	:	2021-09-16
Resolution	:	2.04 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 (1)) were used in the production of this report:

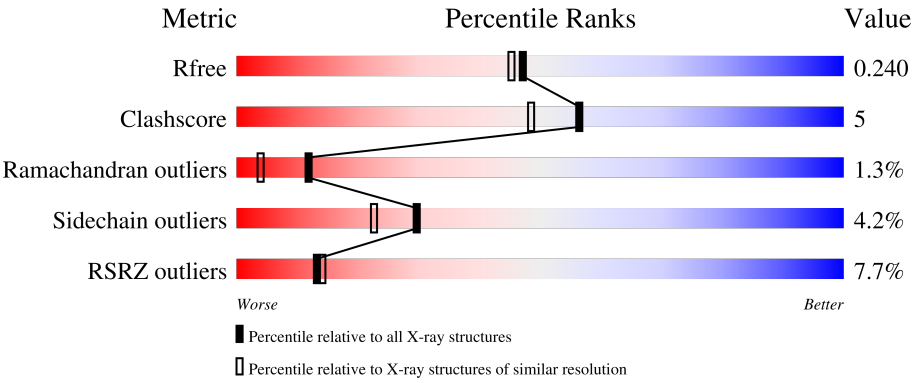
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

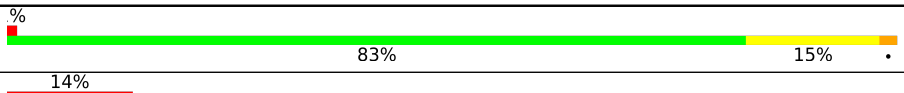
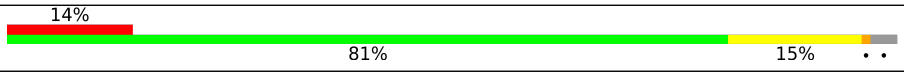
The reported resolution of this entry is 2.04 Å.

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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 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	275	
1	BBB	275	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4474 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 CRISPR-associated protein, APE2256 family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	275	Total	C	N	O	S	0	0	0
			2227	1438	357	430	2			
1	BBB	268	Total	C	N	O	S	0	0	0
			2171	1403	350	416	2			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	12	GLY	SER	engineered mutation	UNP F0NH89
AAA	169	GLY	LYS	engineered mutation	UNP F0NH89
AAA	269	GLY	-	expression tag	UNP F0NH89
AAA	270	SER	-	expression tag	UNP F0NH89
AAA	271	GLU	-	expression tag	UNP F0NH89
AAA	272	PHE	-	expression tag	UNP F0NH89
AAA	273	GLU	-	expression tag	UNP F0NH89
AAA	274	LEU	-	expression tag	UNP F0NH89
AAA	275	GLU	-	expression tag	UNP F0NH89
BBB	12	GLY	SER	engineered mutation	UNP F0NH89
BBB	169	GLY	LYS	engineered mutation	UNP F0NH89
BBB	269	GLY	-	expression tag	UNP F0NH89
BBB	270	SER	-	expression tag	UNP F0NH89
BBB	271	GLU	-	expression tag	UNP F0NH89
BBB	272	PHE	-	expression tag	UNP F0NH89
BBB	273	GLU	-	expression tag	UNP F0NH89
BBB	274	LEU	-	expression tag	UNP F0NH89
BBB	275	GLU	-	expression tag	UNP F0NH89

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	44	Total	O	0	0
			44	44		

Continued on next page...

Continued from previous page...

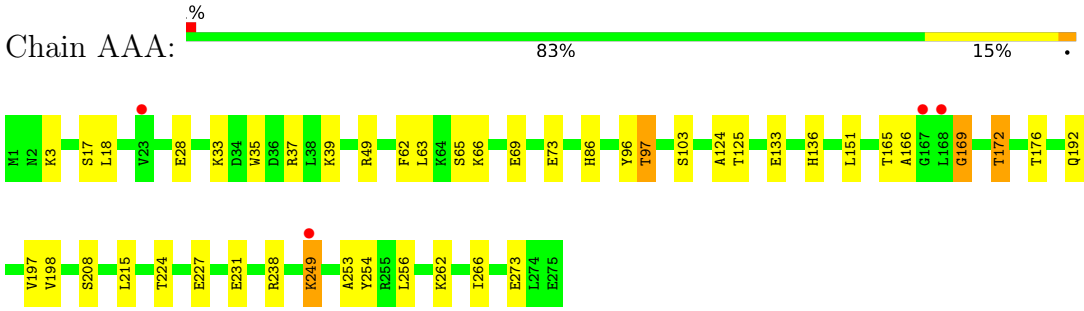
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	BBB	32	Total	O	0	0
			32	32		

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.

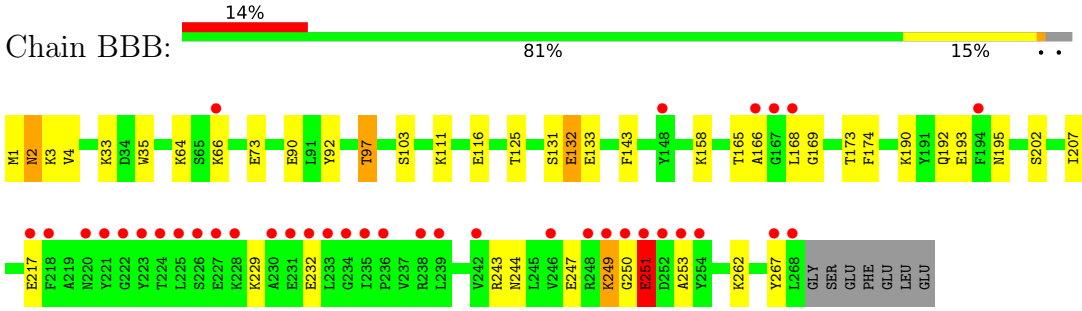
• Molecule 1: CRISPR-associated protein, APE2256 family

Chain AAA:



• Molecule 1: CRISPR-associated protein, APE2256 family

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	76.93Å 104.97Å 78.31Å 90.00° 102.89° 90.00°	Depositor
Resolution (Å)	43.93 – 2.04 43.93 – 2.04	Depositor EDS
% Data completeness (in resolution range)	99.1 (43.93-2.04) 99.1 (43.93-2.04)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.29 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.194 , 0.238 0.200 , 0.240	Depositor DCC
R_{free} test set	1913 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	41.1	Xtriage
Anisotropy	0.307	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 42.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4474	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.32% 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

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	0.88	3/2268 (0.1%)	0.99	3/3058 (0.1%)
1	BBB	0.87	2/2211 (0.1%)	0.94	1/2982 (0.0%)
All	All	0.87	5/4479 (0.1%)	0.96	4/6040 (0.1%)

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	AAA	0	1
1	BBB	0	1
All	All	0	2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	AAA	231	GLU	CD-OE1	-7.33	1.17	1.25
1	BBB	90	GLU	CD-OE1	5.46	1.31	1.25
1	BBB	202	SER	CA-CB	5.17	1.60	1.52
1	AAA	133	GLU	C-O	5.14	1.33	1.23
1	AAA	273	GLU	CD-OE2	5.09	1.31	1.25

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	BBB	97	THR	CA-CB-OG1	-6.39	95.58	109.00
1	AAA	194	PHE	CB-CA-C	-6.17	98.07	110.40
1	AAA	49	ARG	NE-CZ-NH2	-5.81	117.39	120.30
1	AAA	97	THR	CA-CB-OG1	-5.24	98.00	109.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	169	GLY	Peptide
1	BBB	251	GLU	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	AAA	2227	0	2251	21	0
1	BBB	2171	0	2205	25	0
2	AAA	44	0	0	1	0
2	BBB	32	0	0	1	0
All	All	4474	0	4456	44	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 5.

The worst 5 of 44 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:BBB:250:GLY:HA3	1:BBB:253:ALA:HB3	1.44	0.99
1:BBB:1:MET:O	1:BBB:158:LYS:HE3	1.63	0.98
1:AAA:73:GLU:HG3	1:AAA:166:ALA:HB2	1.61	0.82
1:AAA:136:HIS:HD2	2:AAA:339:HOH:O	1.62	0.81
1:BBB:73:GLU:HG3	1:BBB:166:ALA:HB2	1.66	0.78

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	AAA	273/275 (99%)	264 (97%)	7 (3%)	2 (1%)	22 12
1	BBB	266/275 (97%)	251 (94%)	10 (4%)	5 (2%)	8 2
All	All	539/550 (98%)	515 (96%)	17 (3%)	7 (1%)	12 4

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	165	THR
1	BBB	165	THR
1	BBB	169	GLY
1	BBB	251	GLU
1	BBB	193	GLU

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	243/243 (100%)	232 (96%)	11 (4%)	27 20
1	BBB	237/243 (98%)	228 (96%)	9 (4%)	33 26
All	All	480/486 (99%)	460 (96%)	20 (4%)	30 22

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	66	LYS
1	BBB	217	GLU
1	BBB	262	LYS
1	BBB	251	GLU
1	AAA	172	THR

Sometimes 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 monosaccharides 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	AAA	275/275 (100%)	0.02	4 (1%) 73 76	34, 49, 85, 121	0
1	BBB	268/275 (97%)	0.75	38 (14%) 2 2	28, 48, 132, 157	0
All	All	543/550 (98%)	0.38	42 (7%) 13 14	28, 49, 117, 157	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	239	LEU	11.2
1	BBB	230	ALA	10.9
1	BBB	221	TYR	10.5
1	BBB	234	GLY	8.7
1	AAA	168	LEU	8.6

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.