



wwPDB X-ray Structure Validation Summary Report i

Aug 25, 2014 – 05:44 PM EDT

PDB ID : 4MJK
Title : Crystal structure of a CRISPR protein from *Archaeoglobus fulgidus*
Authors : Huang, Q.Q.
Deposited on : 2013-09-03
Resolution : 1.95 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

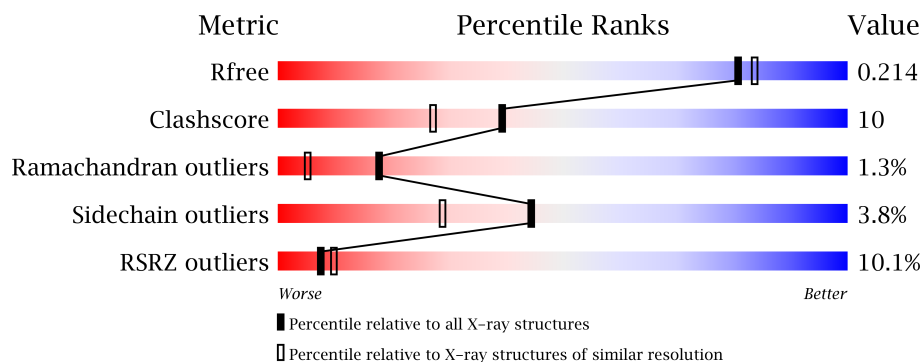
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.16 November 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable23489
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable23489

1 Overall quality at a glance

The reported resolution of this entry is 1.95 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}	66092	1321 (1.96-1.96)
Clashscore	79885	1488 (1.96-1.96)
Ramachandran outliers	78287	1475 (1.96-1.96)
Sidechain outliers	78261	1475 (1.96-1.96)
RSRZ outliers	66119	1321 (1.96-1.96)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	265	
1	B	265	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3953 atoms, of which 0 are hydrogen and 0 are deuterium.

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 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	239	Total	C	N	O	S	Se	0	0	0
			1871	1216	306	343	3	3			
1	B	235	Total	C	N	O	S	Se	0	0	0
			1850	1199	306	340	3	2			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	112	Total	O	0	0
			112	112		
2	B	120	Total	O	0	0
			120	120		

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	42.25Å 87.52Å 140.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.47 – 1.95 43.76 – 1.95	Depositor EDS
% Data completeness (in resolution range)	94.1 (33.47-1.95) 93.3 (43.76-1.95)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.48 (at 1.95Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.195 , 0.216 0.191 , 0.214	Depositor DCC
R_{free} test set	1926 reflections (5.26%)	DCC
Wilson B-factor (Å ²)	19.2	Xtriage
Anisotropy	0.131	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 45.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 38133 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3953	wwPDB-VP
Average B, all atoms (Å ²)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 19.01% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.41	0/1909	0.55	0/2576
1	B	0.40	0/1888	0.53	0/2548
All	All	0.41	0/3797	0.54	0/5124

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1871	0	1900	36	0
1	B	1850	0	1882	38	0
2	A	112	0	0	5	0
2	B	120	0	0	8	0
All	All	3953	0	3782	72	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 10.

The worst 5 of 72 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:56:LYS:HA	1:A:58:ILE:H	1.32	0.95

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:189:ARG:HG2	1:B:245:GLU:HG3	1.51	0.93
1:B:142:LYS:HG3	1:B:143:ILE:HD12	1.61	0.82
1:A:217:MSE:N	1:A:217:MSE:HE2	1.97	0.79
1:A:264:CYS:O	2:A:401:HOH:O	2.00	0.79

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	231/265 (87%)	215 (93%)	11 (5%)	5 (2%)	10	2
1	B	229/265 (86%)	220 (96%)	8 (4%)	1 (0%)	43	30
All	All	460/530 (87%)	435 (95%)	19 (4%)	6 (1%)	18	5

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	221	SER
1	B	89	ALA
1	A	47	ASP
1	A	48	LYS
1	A	195	ARG

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	200/223 (90%)	192 (96%)	8 (4%)	42	26
1	B	198/223 (89%)	191 (96%)	7 (4%)	48	32
All	All	398/446 (89%)	383 (96%)	15 (4%)	44	29

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	223	LYS
1	A	239	THR
1	B	223	LYS
1	A	221	SER
1	B	184	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	65	ASN
1	B	53	GLN
1	B	169	GLN

5.3.3 RNA ⓘ

There are no RNA chains 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, 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(Å ²)	Q<0.9
1	A	239/265 (90%)	0.84	27 (11%) 6 6	9, 20, 46, 64	0
1	B	235/265 (88%)	0.70	21 (8%) 10 12	9, 22, 41, 51	0
All	All	474/530 (89%)	0.77	48 (10%) 7 10	9, 21, 44, 64	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	50	ILE	8.4
1	A	207	VAL	8.0
1	A	104	LEU	7.9
1	A	208	LYS	6.3
1	A	49	GLY	5.2

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 carbohydrates 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.