



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 06:51 PM GMT

PDB ID : 4N06  
Title : Crystal structure of Cas1 from *Archaeoglobus fulgidus* and its nucleolytic activity  
Authors : Kim, T.Y.; Shin, M.; Yen, L.H.T.; Kim, J.S.  
Deposited on : 2013-10-01  
Resolution : 2.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

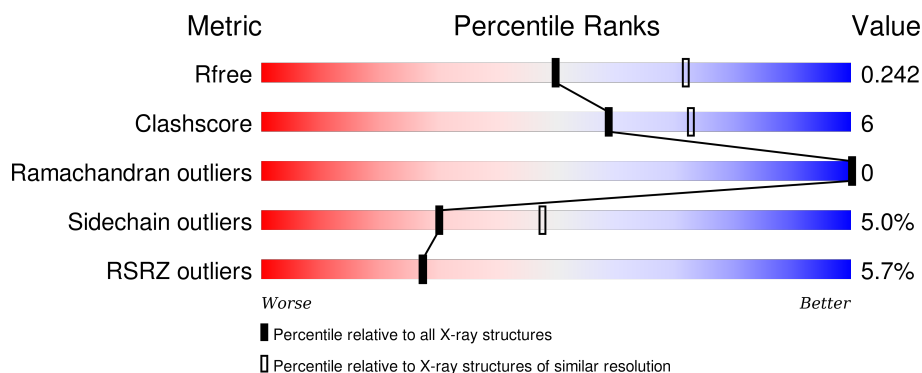
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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))
$R_{free}$	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . 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	347	<div> <div>4%</div> <div>83%</div> <div>14%</div> <div>•</div> </div>
1	B	347	<div> <div>7%</div> <div>84%</div> <div>15%</div> <div>•</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5846 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 endonuclease Cas1 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	346	Total	C	N	O	S	Se	0	0	0
			2739	1736	498	494	4	7			
1	B	347	Total	C	N	O	S	Se	0	0	0
			2752	1747	500	494	4	7			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP O28401
A	0	GLY	-	EXPRESSION TAG	UNP O28401
B	-1	GLY	-	EXPRESSION TAG	UNP O28401
B	0	GLY	-	EXPRESSION TAG	UNP O28401

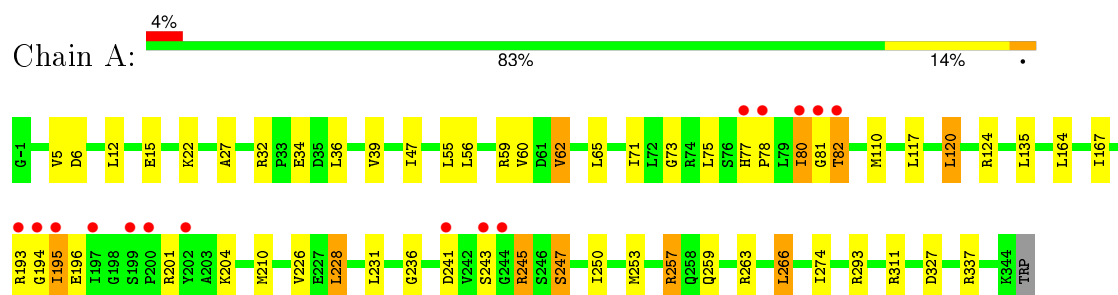
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	225	Total	O	0	0
			225	225		
2	B	130	Total	O	0	0
			130	130		

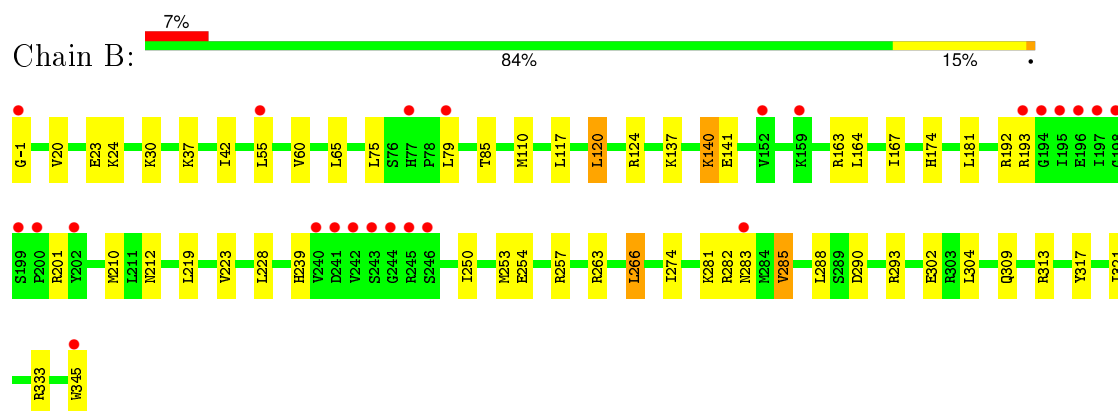
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ( $\text{RSRZ} > 2$ ). 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 endonuclease Cas1 1



- Molecule 1: CRISPR-associated endonuclease Cas1 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.89Å 41.93Å 125.29Å 90.00° 98.03° 90.00°	Depositor
Resolution (Å)	65.12 – 2.40 65.11 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.4 (65.12-2.40) 99.4 (65.11-2.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.61 (at 2.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, $R_{free}$	0.199 , 0.241 0.196 , 0.242	Depositor DCC
$R_{free}$ test set	1736 reflections (6.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	25.6	Xtriage
Anisotropy	0.288	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 47.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 28946 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	5846	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.54% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	A	0.25	0/2767	0.43	0/3701
1	B	0.23	0/2782	0.41	0/3722
All	All	0.24	0/5549	0.42	0/7423

There are no bond length outliers.

There are no bond angle outliers.

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	2739	0	2863	35	0
1	B	2752	0	2873	31	0
2	A	225	0	0	3	0
2	B	130	0	0	4	0
All	All	5846	0	5736	65	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 6.

All (65) 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:194:GLY:HA2	1:A:201:ARG:H	1.51	0.76

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:195:ILE:HG12	1:A:257:ARG:HH22	1.60	0.67
1:A:62:VAL:HG13	1:A:75:LEU:HB2	1.77	0.67
1:B:263:ARG:HH12	1:B:345:TRP:HB3	1.60	0.66
1:B:110:MSE:HE1	1:B:167:ILE:HG22	1.78	0.64
1:B:281:LYS:NZ	1:B:283:ASN:O	2.33	0.62
1:B:23:GLU:HG2	1:B:24:LYS:HG3	1.81	0.62
1:A:110:MSE:HE1	1:A:167:ILE:HG22	1.82	0.60
1:B:124:ARG:NE	2:B:401:HOH:O	2.26	0.60
1:B:239:HIS:NE2	1:B:254:GLU:OE1	2.31	0.59
1:B:210:MSE:HE3	1:B:274:ILE:HD12	1.85	0.59
1:B:290:ASP:OD1	1:B:293:ARG:NH2	2.37	0.56
1:A:259:GLN:HE22	1:A:263:ARG:HH11	1.54	0.56
1:A:311:ARG:NH2	1:A:327:ASP:OD2	2.35	0.55
1:A:124:ARG:NE	2:A:606:HOH:O	2.26	0.54
1:A:22:LYS:HG2	1:A:27:ALA:HA	1.89	0.54
1:A:6:ASP:OD1	1:A:293:ARG:NH2	2.40	0.53
1:A:245:ARG:HG2	1:A:250:ILE:HD11	1.91	0.53
1:A:236:GLY:HA3	1:A:247:SER:HB2	1.90	0.53
1:B:333:ARG:NH2	2:B:423:HOH:O	2.42	0.53
1:A:124:ARG:NH1	2:A:610:HOH:O	2.40	0.53
1:B:192:ARG:NH1	1:B:212:ASN:OD1	2.41	0.53
1:A:196:GLU:HG2	1:A:245:ARG:HH11	1.74	0.53
1:B:250:ILE:HA	1:B:253:MSE:HE2	1.91	0.53
1:A:226:VAL:HG13	1:A:231:LEU:HB2	1.91	0.52
1:A:82:THR:O	1:A:82:THR:OG1	2.28	0.52
1:A:253:MSE:HG2	1:A:257:ARG:HD2	1.92	0.51
1:B:20:VAL:HG22	1:B:30:LYS:HG2	1.93	0.51
1:B:55:LEU:HD22	1:B:60:VAL:HG11	1.91	0.51
1:B:333:ARG:NE	2:B:503:HOH:O	2.41	0.50
1:A:120:LEU:HB3	1:A:135:LEU:HD13	1.94	0.49
1:A:193:ARG:HA	1:A:201:ARG:HH21	1.77	0.48
1:A:65:LEU:HD23	1:A:71:ILE:HA	1.95	0.48
1:B:137:LYS:HA	1:B:137:LYS:HD3	1.73	0.47
1:A:73:GLY:HA3	1:B:75:LEU:HD22	1.97	0.46
1:A:80:ILE:HA	1:A:81:GLY:HA2	1.66	0.46
1:A:77:HIS:CG	1:A:78:PRO:HD2	2.51	0.45
1:A:210:MSE:HE3	1:A:274:ILE:HD12	1.99	0.45
1:A:5:VAL:HG21	1:A:12:LEU:HD13	1.99	0.45
1:A:39:VAL:HG21	1:A:55:LEU:HD13	1.98	0.44
1:B:193:ARG:O	1:B:201:ARG:HB2	2.18	0.44
1:A:195:ILE:HD12	1:A:195:ILE:HA	1.75	0.43

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:253:MSE:O	1:B:257:ARG:HG3	2.18	0.43
1:A:32:ARG:HG3	1:A:34:GLU:HG2	2.01	0.43
1:B:317:TYR:O	1:B:321:ILE:HG13	2.18	0.43
1:B:163:ARG:O	1:B:167:ILE:HG13	2.19	0.43
1:A:266:LEU:HA	1:A:266:LEU:HD12	1.89	0.43
1:B:219:LEU:O	1:B:223:VAL:HG23	2.19	0.43
1:B:120:LEU:HD12	1:B:120:LEU:HA	1.88	0.42
1:A:228:LEU:HA	1:A:228:LEU:HD13	1.88	0.42
1:B:141:GLU:OE1	1:B:174:HIS:NE2	2.53	0.42
1:A:12:LEU:HB3	1:A:47:ILE:HG13	2.02	0.41
1:B:85:THR:HG22	1:B:333:ARG:HH11	1.84	0.41
1:B:313:ARG:NH2	2:B:469:HOH:O	2.44	0.41
1:A:250:ILE:HA	1:A:253:MSE:HE2	2.01	0.41
1:B:282:ARG:O	1:B:285:VAL:HG12	2.19	0.41
1:B:-1:GLY:HA3	1:B:37:LYS:HB2	2.03	0.41
1:B:137:LYS:HD2	1:B:140:LYS:HD2	2.01	0.41
1:B:42:ILE:HG12	1:B:65:LEU:HB2	2.01	0.41
1:A:253:MSE:O	1:A:257:ARG:HG2	2.21	0.41
1:A:55:LEU:HD22	1:A:60:VAL:HG11	2.03	0.41
1:B:266:LEU:HA	1:B:266:LEU:HD12	1.96	0.40
1:A:337:ARG:NH2	2:A:614:HOH:O	2.54	0.40
1:A:241:ASP:OD1	1:A:243:SER:HB3	2.21	0.40
1:B:228:LEU:HA	1:B:228:LEU:HD23	1.98	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	344/347 (99%)	337 (98%)	7 (2%)	0	100	100
1	B	345/347 (99%)	340 (99%)	5 (1%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	689/694 (99%)	677 (98%)	12 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	290/284 (102%)	273 (94%)	17 (6%)	24	38
1	B	291/284 (102%)	279 (96%)	12 (4%)	37	57
All	All	581/568 (102%)	552 (95%)	29 (5%)	30	48

All (29) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	15	GLU
1	A	36	LEU
1	A	56	LEU
1	A	59	ARG
1	A	62	VAL
1	A	80	ILE
1	A	82	THR
1	A	117	LEU
1	A	120	LEU
1	A	164	LEU
1	A	195	ILE
1	A	204	LYS
1	A	228	LEU
1	A	245	ARG
1	A	247	SER
1	A	257	ARG
1	A	266	LEU
1	B	79	LEU
1	B	117	LEU
1	B	120	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	140	LYS
1	B	164	LEU
1	B	181	LEU
1	B	266	LEU
1	B	285	VAL
1	B	288	LEU
1	B	302	GLU
1	B	304	LEU
1	B	309	GLN

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

Mol	Chain	Res	Type
1	A	119	ASN
1	A	259	GLN

### 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 ⓘ

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	339/347 (97%)	-0.28	15 (4%) 38 39	4, 15, 74, 110	0
1	B	340/347 (97%)	0.15	24 (7%) 19 19	11, 31, 79, 121	0
All	All	679/694 (97%)	-0.06	39 (5%) 27 27	4, 23, 76, 121	0

All (39) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	198	GLY	10.2
1	B	244	GLY	8.3
1	B	242	VAL	6.3
1	B	245	ARG	5.8
1	B	197	ILE	5.3
1	A	244	GLY	4.9
1	B	195	ILE	4.9
1	B	202	TYR	4.3
1	A	197	ILE	4.3
1	A	243	SER	4.2
1	A	82	THR	4.0
1	A	78	PRO	3.8
1	B	240	VAL	3.8
1	A	199	SER	3.7
1	B	200	PRO	3.6
1	B	246	SER	3.6
1	B	345	TRP	3.6
1	B	159	LYS	3.4
1	B	193	ARG	3.4
1	A	80	ILE	3.0
1	A	202	TYR	2.9
1	A	193	ARG	2.8
1	B	-1	GLY	2.8
1	A	194	GLY	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	241	ASP	2.6
1	B	77	HIS	2.6
1	B	79	LEU	2.5
1	B	196	GLU	2.3
1	B	194	GLY	2.3
1	A	195	ILE	2.2
1	A	77	HIS	2.2
1	B	152	VAL	2.2
1	B	199	SER	2.2
1	B	283	ASN	2.2
1	A	200	PRO	2.2
1	A	81	GLY	2.2
1	B	55	LEU	2.1
1	B	243	SER	2.1
1	A	241	ASP	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.