



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 04:45 AM GMT

PDB ID : 2O3A  
Title : Crystal structure of a protein AF\_0751 from Archaeoglobus fulgidus  
Authors : Bonanno, J.B.; Dickey, M.; Bain, K.T.; Wu, B.; Slocombe, A.; Sridhar, V.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2006-12-01  
Resolution : 2.20 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

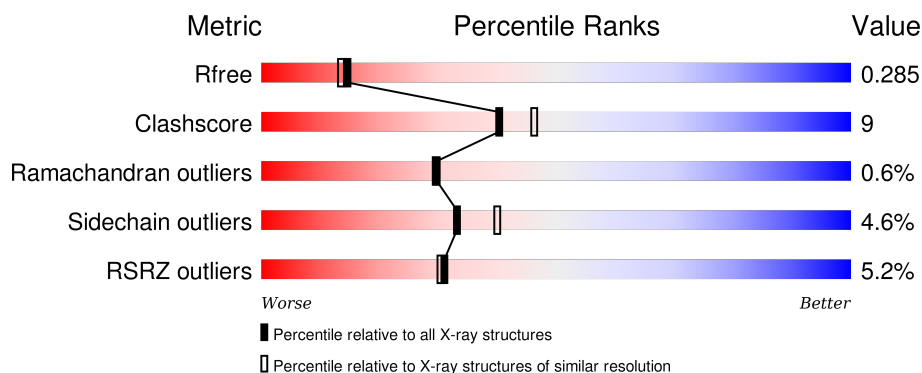
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

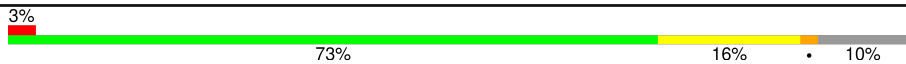
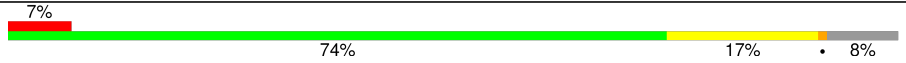
The reported resolution of this entry is 2.20 Å.

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	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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	178	
1	B	178	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2679 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 UPF0106 protein AF\_0751.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	161	Total	C	N	O	S	0	0	0
			1284	833	221	228	2			
1	B	163	Total	C	N	O	S	0	0	0
			1304	844	226	232	2			

There are 22 discrepancies between the modelled and reference sequences:

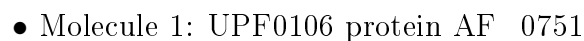
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	CLONING ARTIFACT	UNP O29507
A	0	SER	-	CLONING ARTIFACT	UNP O29507
A	1	LEU	-	CLONING ARTIFACT	UNP O29507
A	169	GLU	-	CLONING ARTIFACT	UNP O29507
A	170	GLY	-	CLONING ARTIFACT	UNP O29507
A	171	HIS	-	EXPRESSION TAG	UNP O29507
A	172	HIS	-	EXPRESSION TAG	UNP O29507
A	173	HIS	-	EXPRESSION TAG	UNP O29507
A	174	HIS	-	EXPRESSION TAG	UNP O29507
A	175	HIS	-	EXPRESSION TAG	UNP O29507
A	176	HIS	-	EXPRESSION TAG	UNP O29507
B	-1	MET	-	CLONING ARTIFACT	UNP O29507
B	0	SER	-	CLONING ARTIFACT	UNP O29507
B	1	LEU	-	CLONING ARTIFACT	UNP O29507
B	169	GLU	-	CLONING ARTIFACT	UNP O29507
B	170	GLY	-	CLONING ARTIFACT	UNP O29507
B	171	HIS	-	EXPRESSION TAG	UNP O29507
B	172	HIS	-	EXPRESSION TAG	UNP O29507
B	173	HIS	-	EXPRESSION TAG	UNP O29507
B	174	HIS	-	EXPRESSION TAG	UNP O29507
B	175	HIS	-	EXPRESSION TAG	UNP O29507
B	176	HIS	-	EXPRESSION TAG	UNP O29507

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	41	Total 41	O 41	0	0
2	B	50	Total 50	O 50	0	0



- Molecule 1: UPF0106 protein AF 0751



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	53.29 Å 55.89 Å 108.78 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.04 – 2.20 19.03 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.7 (19.04-2.20) 97.7 (19.03-2.20)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.83 (at 2.21 Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.230 , 0.292 0.225 , 0.285	Depositor DCC
$R_{free}$ test set	851 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	35.4	Xtriage
Anisotropy	0.893	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 49.8	EDS
Estimated twinning fraction	0.043 for k,h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	0 of 16684 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2679	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.23% 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 [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	0.52	0/1309	0.68	0/1764
1	B	0.51	0/1330	0.66	1/1793 (0.1%)
All	All	0.51	0/2639	0.67	1/3557 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	40	ASP	N-CA-C	6.12	127.52	111.00

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	1284	0	1328	30	0
1	B	1304	0	1348	26	0
2	A	41	0	0	0	0
2	B	50	0	0	2	0
All	All	2679	0	2676	46	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 9.

All (46) 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:14:ARG:HH11	1:B:14:ARG:HG2	1.36	0.91
1:A:72:PHE:HA	1:A:73:ASP:CB	2.07	0.84
1:A:155:LYS:HB2	1:A:166:VAL:HG13	1.61	0.81
1:A:72:PHE:HA	1:A:73:ASP:HB2	1.67	0.76
1:A:27:ARG:HH12	1:B:125:GLN:HE21	1.36	0.74
1:A:21:HIS:HD2	1:B:128:SER:OG	1.75	0.69
1:A:21:HIS:CD2	1:B:128:SER:OG	2.47	0.68
1:B:14:ARG:NH1	1:B:14:ARG:HG2	2.09	0.66
1:A:12:PRO:HB3	1:A:16:LYS:NZ	2.15	0.62
1:A:1:LEU:HD11	1:A:99:VAL:HG23	1.83	0.61
1:A:72:PHE:HA	1:A:73:ASP:HB3	1.83	0.60
1:A:27:ARG:NH1	1:B:125:GLN:HE21	2.01	0.58
1:B:2:GLU:HG3	1:B:98:LYS:HG2	1.84	0.58
1:A:128:SER:OG	1:B:21:HIS:HD2	1.88	0.56
1:A:12:PRO:HB3	1:A:16:LYS:HZ3	1.71	0.55
1:A:2:GLU:HG2	1:A:32:LYS:HE2	1.89	0.54
1:B:57:ASP:N	1:B:57:ASP:OD1	2.41	0.52
1:B:67:LYS:O	1:B:71:GLU:HG2	2.10	0.52
1:A:72:PHE:CA	1:A:73:ASP:CB	2.87	0.51
1:A:27:ARG:HH22	1:B:125:GLN:NE2	2.07	0.51
1:A:67:LYS:O	1:A:71:GLU:HG2	2.13	0.49
1:B:157:ILE:HD11	1:B:166:VAL:CG2	2.42	0.49
1:A:128:SER:OG	1:B:21:HIS:CD2	2.66	0.49
1:B:77:VAL:O	1:B:101:VAL:HA	2.13	0.48
1:A:27:ARG:HH22	1:B:125:GLN:HE22	1.62	0.48
1:A:72:PHE:CA	1:A:73:ASP:HB2	2.42	0.48
1:A:6:LEU:HD13	1:A:68:LEU:HD23	1.97	0.46
1:A:1:LEU:HD11	1:A:99:VAL:CG2	2.44	0.46
1:A:84:ILE:HG21	1:A:89:LYS:HE2	1.97	0.45
1:B:127:HIS:CG	1:B:128:SER:H	2.35	0.45
1:B:14:ARG:NH1	1:B:14:ARG:CG	2.79	0.45
1:B:79:LEU:HG	1:B:101:VAL:HG23	1.98	0.45
1:A:11:ARG:O	1:A:15:ASP:HB2	2.17	0.44
1:A:1:LEU:CD1	1:A:99:VAL:HG23	2.47	0.43
1:A:69:LEU:HD12	1:A:69:LEU:HA	1.88	0.43
1:B:1:LEU:HD11	1:B:99:VAL:HG23	2.01	0.43
1:B:10:HIS:HB3	2:B:214:HOH:O	2.18	0.43
1:B:84:ILE:HG21	1:B:89:LYS:HE2	1.99	0.42
1:A:27:ARG:HH12	1:B:125:GLN:NE2	2.11	0.42
1:A:10:HIS:ND1	1:A:15:ASP:HB3	2.34	0.42
1:A:3:VAL:HG11	1:A:137:LEU:CD2	2.50	0.42
1:B:109:PRO:HA	1:B:110:PRO:HD2	1.87	0.41

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:156:VAL:HB	1:B:124:THR:HB	2.02	0.41
1:A:11:ARG:HA	1:A:12:PRO:HD3	1.80	0.41
1:B:64:SER:HA	2:B:181:HOH:O	2.20	0.40
1:B:6:LEU:HD22	1:B:68:LEU:HD23	2.02	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	155/178 (87%)	147 (95%)	6 (4%)	2 (1%)	15	11
1	B	159/178 (89%)	152 (96%)	7 (4%)	0	100	100
All	All	314/356 (88%)	299 (95%)	13 (4%)	2 (1%)	30	29

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	73	ASP
1	A	127	HIS

### 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.

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	140/155 (90%)	134 (96%)	6 (4%)	35	43
1	B	142/155 (92%)	135 (95%)	7 (5%)	31	36
All	All	282/310 (91%)	269 (95%)	13 (5%)	33	40

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	ARG
1	A	17	ARG
1	A	42	SER
1	A	69	LEU
1	A	98	LYS
1	A	166	VAL
1	B	7	ARG
1	B	40	ASP
1	B	66	LYS
1	B	101	VAL
1	B	146	ASP
1	B	147	ILE
1	B	159	SER

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	21	HIS
1	B	21	HIS
1	B	125	GLN

### 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 [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 ⓘ

### 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	161/178 (90%)	0.28	5 (3%) 52 51	29, 46, 57, 60	1 (0%)
1	B	163/178 (91%)	0.40	12 (7%) 17 17	27, 45, 58, 61	2 (1%)
All	All	324/356 (91%)	0.34	17 (5%) 31 30	27, 45, 58, 61	3 (0%)

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	14	ARG	4.1
1	B	101	VAL	3.6
1	A	11	ARG	3.4
1	B	146	ASP	3.4
1	B	102	VAL	3.3
1	B	6	LEU	3.1
1	A	5	VAL	2.9
1	B	110	PRO	2.9
1	B	77	VAL	2.8
1	B	39	GLU	2.7
1	A	77	VAL	2.6
1	A	53	ARG	2.2
1	A	39	GLU	2.2
1	B	114	GLU	2.1
1	B	65	TRP	2.0
1	B	5	VAL	2.0
1	B	70	ARG	2.0

### 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 [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.