



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2020 – 12:20 PM BST

PDB ID : 5WU7  
Title : Crystal structure of GH57-type branching enzyme from hyperthermophilic archaeon *Pyrococcus horikoshii*  
Authors : Na, S.; Jo, I.; Ha, N.-C.  
Deposited on : 2016-12-16  
Resolution : 2.31 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](mailto: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 ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.13.1  
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) : 2.13.1

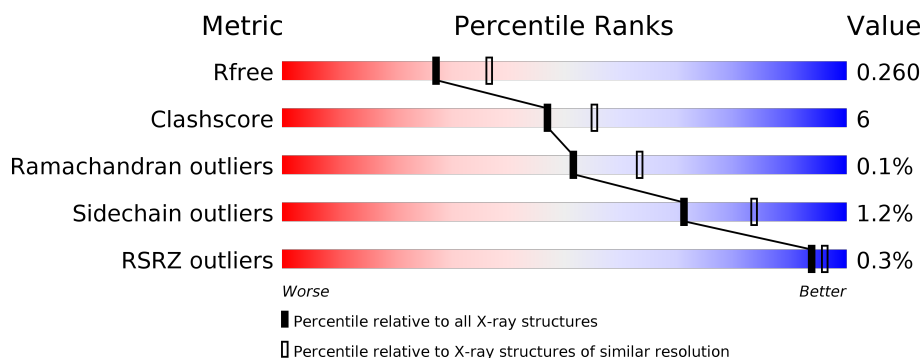
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.31 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	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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 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  $\leq 5\%$ . 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	568	
1	B	568	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9009 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 Uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	530	Total	C	N	O	S	0	0	0
			4430	2890	739	788	13			
1	B	536	Total	C	N	O	S	0	0	0
			4471	2913	747	798	13			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	561	LEU	-	expression tag	UNP O50094
A	562	GLU	-	expression tag	UNP O50094
A	563	HIS	-	expression tag	UNP O50094
A	564	HIS	-	expression tag	UNP O50094
A	565	HIS	-	expression tag	UNP O50094
A	566	HIS	-	expression tag	UNP O50094
A	567	HIS	-	expression tag	UNP O50094
A	568	HIS	-	expression tag	UNP O50094
B	561	LEU	-	expression tag	UNP O50094
B	562	GLU	-	expression tag	UNP O50094
B	563	HIS	-	expression tag	UNP O50094
B	564	HIS	-	expression tag	UNP O50094
B	565	HIS	-	expression tag	UNP O50094
B	566	HIS	-	expression tag	UNP O50094
B	567	HIS	-	expression tag	UNP O50094
B	568	HIS	-	expression tag	UNP O50094

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	57	Total	O	0	0
			57	57		
3	B	39	Total	O	0	0
			39	39		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	120.86Å 42.33Å 122.23Å 90.00° 90.96° 90.00°	Depositor
Resolution (Å)	33.52 – 2.31 33.52 – 2.31	Depositor EDS
% Data completeness (in resolution range)	79.7 (33.52-2.31) 79.7 (33.52-2.31)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.14 (at 2.31Å)	Xtriage
Refinement program	PHENIX (1.11_2567)	Depositor
R, $R_{free}$	0.200 , 0.260 0.200 , 0.260	Depositor DCC
$R_{free}$ test set	2249 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.1	Xtriage
Anisotropy	0.346	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 13.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.084 for l,k,-h 0.038 for h,-k,-l 0.036 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	9009	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.74% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL

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	0/4565	0.41	0/6167
1	B	0.26	0/4607	0.42	0/6224
All	All	0.26	0/9172	0.42	0/12391

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 [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	4430	0	4331	41	0
1	B	4471	0	4376	68	0
2	A	6	0	8	0	0
2	B	6	0	8	0	0
3	A	57	0	0	1	0
3	B	39	0	0	0	0
All	All	9009	0	8723	109	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.

The worst 5 of 109 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:9:LEU:HB2	1:B:58:ILE:HD12	1.44	0.96
1:B:127:GLY:HA2	1:B:130:ARG:HD2	1.56	0.84
1:B:7:PHE:HB2	1:B:56:LEU:HD21	1.61	0.83
1:B:130:ARG:NH2	1:B:173:TYR:O	2.28	0.65
1:B:334:LEU:HG	1:B:382:ILE:HD11	1.78	0.65

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	522/568 (92%)	503 (96%)	19 (4%)	0	100	100
1	B	532/568 (94%)	505 (95%)	26 (5%)	1 (0%)	47	58
All	All	1054/1136 (93%)	1008 (96%)	45 (4%)	1 (0%)	51	63

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	94	ARG

### 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	462/496 (93%)	459 (99%)	3 (1%)	86	93
1	B	466/496 (94%)	458 (98%)	8 (2%)	60	75
All	All	928/992 (94%)	917 (99%)	11 (1%)	71	83

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	87	SER
1	B	115	TYR
1	B	300	ARG
1	B	56	LEU
1	B	246	ARG

Some 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 ⓘ

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	GOL	B	601	-	5,5,5	0.35	0	5,5,5	0.30	0
2	GOL	A	601	-	5,5,5	0.38	0	5,5,5	0.27	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	B	601	-	-	3/4/4/4	-
2	GOL	A	601	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	601	GOL	C1-C2-C3-O3
2	A	601	GOL	O1-C1-C2-C3
2	B	601	GOL	O2-C2-C3-O3
2	A	601	GOL	O1-C1-C2-O2
2	B	601	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [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, 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	530/568 (93%)	-0.44	0 <b>100</b> <b>100</b>	12, 28, 49, 65	0
1	B	536/568 (94%)	-0.22	3 (0%) <b>89</b> <b>92</b>	15, 37, 59, 71	0
All	All	1066/1136 (93%)	-0.33	3 (0%) <b>94</b> <b>96</b>	12, 31, 56, 71	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	308	GLY	2.5
1	B	504	PHE	2.4
1	B	94	ARG	2.3

### 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 monosaccharides in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	GOL	B	601	6/6	0.89	0.14	36,43,48,52	0
2	GOL	A	601	6/6	0.91	0.10	30,31,33,37	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.