



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 05:08 am GMT

PDB ID : 2IT2  
Title : Structure of PH1069 protein from *Pyrococcus horikoshii*  
Authors : Lokanath, N.K.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2006-10-18  
Resolution : 1.50 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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

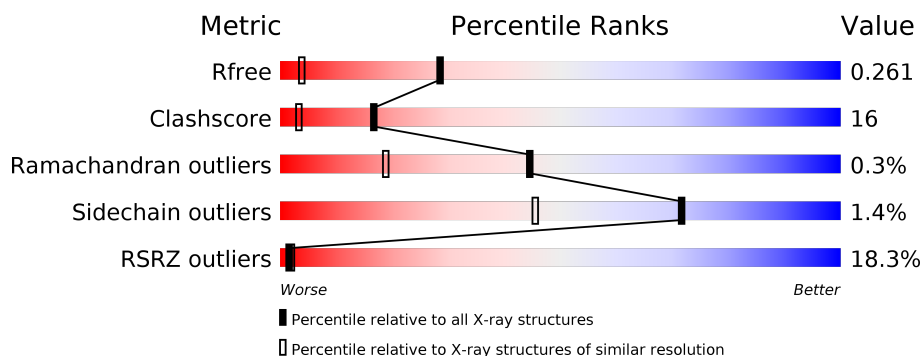
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.50 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))
$R_{free}$	100719	2279 (1.50-1.50)
Clashscore	112137	2503 (1.50-1.50)
Ramachandran outliers	110173	2445 (1.50-1.50)
Sidechain outliers	110143	2443 (1.50-1.50)
RSRZ outliers	101464	2305 (1.50-1.50)

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	200	<div> <div>21%</div> <div>66%</div> <div>29%</div> <div>6%</div> </div>
1	B	200	<div> <div>14%</div> <div>72%</div> <div>26%</div> <div>•</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3446 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 UPF0130 protein PH1069.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	189	Total	C	N	O	S	Se	0	0	0
			1539	985	267	278	2	7			
1	B	196	Total	C	N	O	S	Se	0	0	0
			1593	1020	277	288	2	6			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	EXPRESSION TAG	UNP O58796
A	2	LEU	-	EXPRESSION TAG	UNP O58796
A	3	LEU	-	EXPRESSION TAG	UNP O58796
A	4	TYR	-	EXPRESSION TAG	UNP O58796
B	1	MSE	-	EXPRESSION TAG	UNP O58796
B	2	LEU	-	EXPRESSION TAG	UNP O58796
B	3	LEU	-	EXPRESSION TAG	UNP O58796
B	4	TYR	-	EXPRESSION TAG	UNP O58796

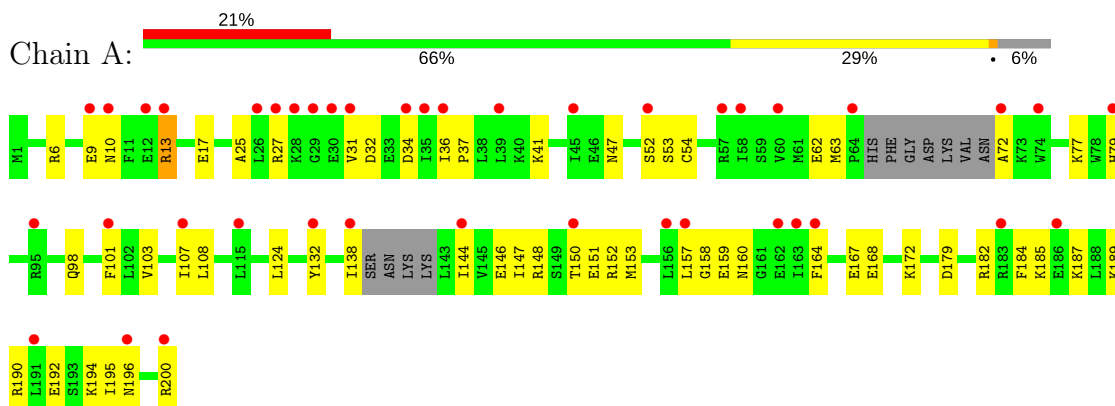
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	151	Total	O	0	0
			151	151		
2	B	163	Total	O	0	0
			163	163		

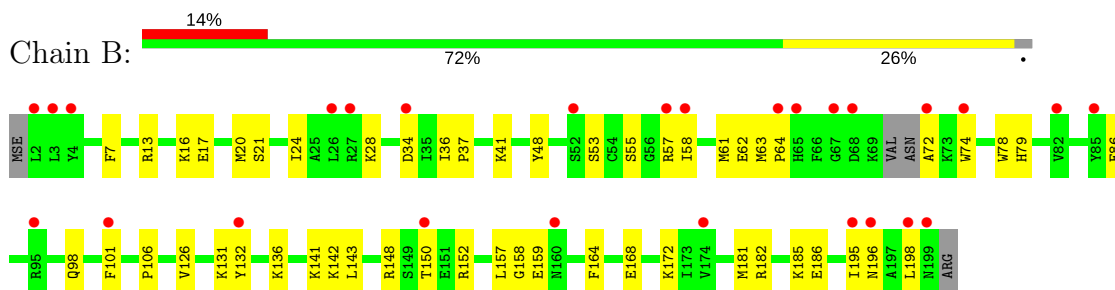
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: UPF0130 protein PH1069



#### • Molecule 1: UPF0130 protein PH1069



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	37.43Å 53.90Å 53.21Å 104.60° 102.90° 109.30°	Depositor
Resolution (Å)	17.55 – 1.50 17.55 – 1.50	Depositor EDS
% Data completeness (in resolution range)	99.1 (17.55-1.50) 83.8 (17.55-1.50)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	10.58 (at 1.50Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.250 , 0.265 0.249 , 0.261	Depositor DCC
$R_{free}$ test set	2703 reflections (5.06%)	DCC
Wilson B-factor (Å <sup>2</sup> )	20.0	Xtriage
Anisotropy	0.314	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.48 , 76.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3446	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.09% 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](#)

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	0/1554	0.66	0/2068
1	B	0.51	2/1611 (0.1%)	0.67	0/2146
All	All	0.46	2/3165 (0.1%)	0.67	0/4214

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	78	TRP	NE1-CE2	8.70	1.48	1.37
1	B	74	TRP	NE1-CE2	8.68	1.48	1.37

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	1539	0	1606	59	0
1	B	1593	0	1655	43	0
2	A	151	0	0	11	0
2	B	163	0	0	12	0
All	All	3446	0	3261	101	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 16.

The worst 5 of 101 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:54:CYS:SG	1:A:107:ILE:HG22	1.93	1.08
1:A:9:GLU:HG2	1:A:10:ASN:HD22	1.36	0.91
1:B:13:ARG:HG2	2:B:307:HOH:O	1.81	0.81
1:A:107:ILE:HD11	1:A:146:GLU:CG	2.11	0.80
1:B:195:ILE:HD12	1:B:196:ASN:N	1.97	0.80

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	183/200 (92%)	176 (96%)	6 (3%)	1 (0%)	32	10
1	B	192/200 (96%)	186 (97%)	6 (3%)	0	100	100
All	All	375/400 (94%)	362 (96%)	12 (3%)	1 (0%)	44	19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	52	SER

### 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	170/173 (98%)	165 (97%)	5 (3%)	48	15
1	B	176/173 (102%)	176 (100%)	0	100	100
All	All	346/346 (100%)	341 (99%)	5 (1%)	71	45

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	6	ARG
1	A	13	ARG
1	A	27	ARG
1	A	32	ASP
1	A	192	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	10	ASN
1	A	47	ASN
1	A	160	ASN
1	B	123	ASN
1	B	196	ASN

### 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 [i](#)

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	182/200 (91%)	1.41	41 (22%) <b>1</b> <b>1</b>	14, 25, 36, 43	0
1	B	190/200 (95%)	1.12	27 (14%) <b>3</b> <b>3</b>	13, 21, 33, 42	0
All	All	372/400 (93%)	1.26	68 (18%) <b>1</b> <b>2</b>	13, 23, 35, 43	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	138	ILE	8.0
1	A	29	GLY	6.4
1	B	85	TYR	5.5
1	A	74	TRP	4.8
1	A	200	ARG	4.8

### 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.