



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 08:10 PM GMT

PDB ID : 1J31  
Title : Crystal Structure of Hypothetical Protein PH0642 from *Pyrococcus horikoshii*  
Authors : Sakai, N.; Tajika, Y.; Yao, M.; Watanabe, N.; Tanaka, I.  
Deposited on : 2003-01-16  
Resolution : 1.60 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

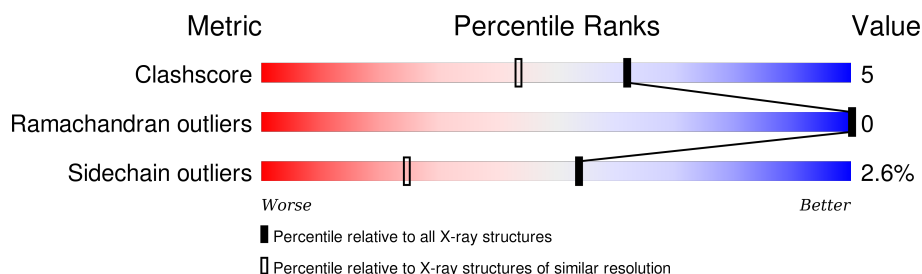
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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))
Clashscore	102246	2732 (1.60-1.60)
Ramachandran outliers	100387	2654 (1.60-1.60)
Sidechain outliers	100360	2653 (1.60-1.60)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	262	 93% 5% •
1	B	262	 91% 7% •
1	C	262	 90% 8% •
1	D	262	 92% 8%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9070 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 Hypothetical protein PH0642.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	262	Total	C	N	O	S	Se	0	0	0
			2103	1359	348	388	1	7			
1	B	262	Total	C	N	O	S	Se	0	0	0
			2103	1359	348	388	1	7			
1	C	262	Total	C	N	O	S	Se	0	0	0
			2103	1359	348	388	1	7			
1	D	262	Total	C	N	O	S	Se	0	0	0
			2103	1359	348	388	1	7			

There are 28 discrepancies between the modelled and reference sequences:

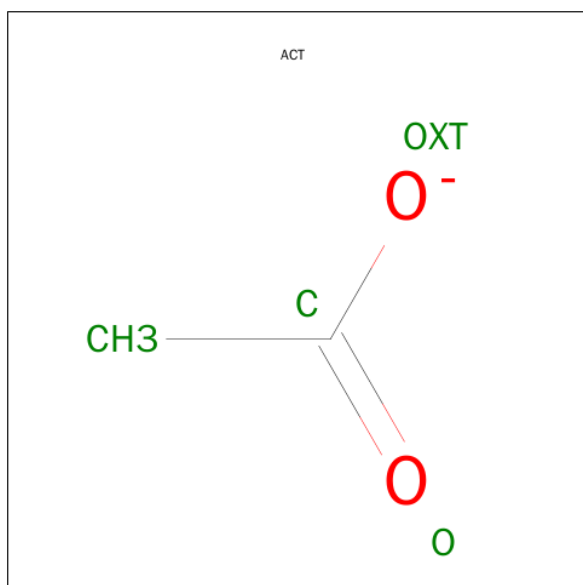
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP O58376
A	9	MSE	MET	MODIFIED RESIDUE	UNP O58376
A	73	MSE	MET	MODIFIED RESIDUE	UNP O58376
A	144	MSE	MET	MODIFIED RESIDUE	UNP O58376
A	174	MSE	MET	MODIFIED RESIDUE	UNP O58376
A	181	MSE	MET	MODIFIED RESIDUE	UNP O58376
A	248	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	1	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	9	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	73	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	144	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	174	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	181	MSE	MET	MODIFIED RESIDUE	UNP O58376
B	248	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	1	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	9	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	73	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	144	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	174	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	181	MSE	MET	MODIFIED RESIDUE	UNP O58376
C	248	MSE	MET	MODIFIED RESIDUE	UNP O58376

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	MSE	MET	MODIFIED RESIDUE	UNP O58376
D	9	MSE	MET	MODIFIED RESIDUE	UNP O58376
D	73	MSE	MET	MODIFIED RESIDUE	UNP O58376
D	144	MSE	MET	MODIFIED RESIDUE	UNP O58376
D	174	MSE	MET	MODIFIED RESIDUE	UNP O58376
D	181	MSE	MET	MODIFIED RESIDUE	UNP O58376
D	248	MSE	MET	MODIFIED RESIDUE	UNP O58376

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula:  $C_2H_3O_2$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0
2	C	1	Total C O 4 2 2	0	0
2	D	1	Total C O 4 2 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	169	Total O 169 169	0	0
3	B	148	Total O 148 148	0	0

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	159	Total	O	0	0
			159	159		
3	D	166	Total	O	0	0
			166	166		

### 3 Residue-property plots

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 ( $RSRZ > 2$ ). 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.

Note EDS was not executed.

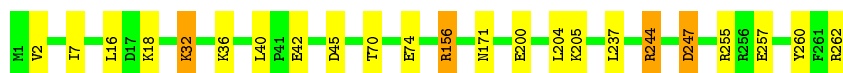
#### • Molecule 1: Hypothetical protein PH0642

Chain A:  93% 5% .




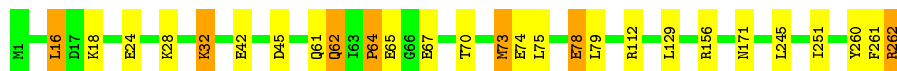
#### • Molecule 1: Hypothetical protein PH0642

Chain B:  91% 7% .



#### • Molecule 1: Hypothetical protein PH0642

Chain C:  90% 8% .



#### • Molecule 1: Hypothetical protein PH0642

Chain D:  92% 8% .



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.04Å 89.02Å 77.87Å 90.00° 96.15° 90.00°	Depositor
Resolution (Å)	10.00 – 1.60	Depositor
% Data completeness (in resolution range)	99.4 (10.00-1.60)	Depositor
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.07	Depositor
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.166 , 0.194	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	9070	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	15.0	wwPDB-VP

## 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: ACT

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	0.68	0/2140	0.85	0/2872
1	B	0.69	0/2140	0.97	6/2872 (0.2%)
1	C	0.67	1/2140 (0.0%)	0.84	0/2872
1	D	0.67	1/2140 (0.0%)	0.82	0/2872
All	All	0.68	2/8560 (0.0%)	0.87	6/11488 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	73	MSE	SE-CE	-8.54	1.45	1.95
1	C	73	MSE	SE-CE	-5.86	1.60	1.95

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	156	ARG	NE-CZ-NH1	17.70	129.15	120.30
1	B	156	ARG	NE-CZ-NH2	-16.49	112.05	120.30
1	B	156	ARG	CD-NE-CZ	7.87	134.62	123.60
1	B	255	ARG	NE-CZ-NH2	-6.75	116.92	120.30
1	B	255	ARG	NE-CZ-NH1	6.34	123.47	120.30

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	2103	0	2124	19	0
1	B	2103	0	2124	21	0
1	C	2103	0	2124	31	0
1	D	2103	0	2124	31	0
2	A	4	0	3	0	0
2	B	4	0	3	0	0
2	C	4	0	3	0	0
2	D	4	0	3	0	0
3	A	169	0	0	1	0
3	B	148	0	0	3	0
3	C	159	0	0	0	0
3	D	166	0	0	3	0
All	All	9070	0	8508	92	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 5.

The worst 5 of 92 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:53:ARG:NH1	1:A:95:LEU:HD22	1.66	1.10
1:A:53:ARG:HH12	1:A:95:LEU:HD22	1.14	1.07
1:D:2:VAL:HG23	1:D:36:LYS:HD3	1.47	0.95
1:C:18:LYS:HB2	1:C:18:LYS:NZ	1.84	0.93
1:A:153:GLU:OE2	1:B:156:ARG:HD3	1.69	0.92

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	260/262 (99%)	248 (95%)	12 (5%)	0	100	100
1	B	260/262 (99%)	252 (97%)	8 (3%)	0	100	100
1	C	260/262 (99%)	250 (96%)	10 (4%)	0	100	100
1	D	260/262 (99%)	253 (97%)	7 (3%)	0	100	100
All	All	1040/1048 (99%)	1003 (96%)	37 (4%)	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	221/214 (103%)	215 (97%)	6 (3%)	52	23
1	B	221/214 (103%)	215 (97%)	6 (3%)	52	23
1	C	221/214 (103%)	214 (97%)	7 (3%)	46	18
1	D	221/214 (103%)	217 (98%)	4 (2%)	66	41
All	All	884/856 (103%)	861 (97%)	23 (3%)	54	25

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	247	ASP
1	C	16	LEU
1	D	204	LEU
1	B	262	ARG
1	C	32	LYS

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

Mol	Chain	Res	Type
1	B	93	ASN
1	D	93	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

4 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	ACT	A	1301	-	1,3,3	3.41	1 (100%)	0,3,3	0.00	-
2	ACT	B	2301	-	1,3,3	2.64	1 (100%)	0,3,3	0.00	-
2	ACT	C	3301	-	1,3,3	3.40	1 (100%)	0,3,3	0.00	-
2	ACT	D	4301	-	1,3,3	3.62	1 (100%)	0,3,3	0.00	-

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	ACT	A	1301	-	-	0/0/0/0	0/0/0/0
2	ACT	B	2301	-	-	0/0/0/0	0/0/0/0
2	ACT	C	3301	-	-	0/0/0/0	0/0/0/0
2	ACT	D	4301	-	-	0/0/0/0	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2301	ACT	CH3-C	2.64	1.52	1.48
2	C	3301	ACT	CH3-C	3.40	1.53	1.48
2	A	1301	ACT	CH3-C	3.41	1.53	1.48
2	D	4301	ACT	CH3-C	3.62	1.53	1.48

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 ⓘ

EDS was not executed - this section will therefore be empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.