wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 10:56 AM EDT

PDB ID	:	2CSU
Title	:	Crystal structure of PH0766 from Pyrococcus horikoshii OT3
Authors	:	Sugahara, M.; Kunishima, N.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-05-23
Resolution	:	2.20 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

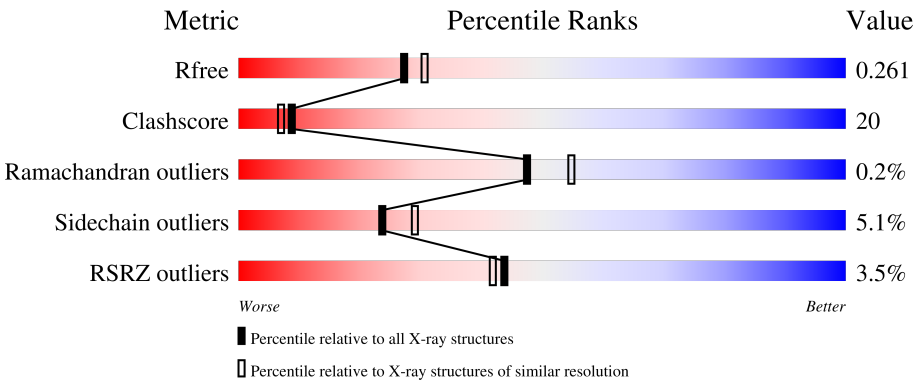
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

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}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	457	% 70% 22% • 5%
1	B	457	6% 56% 35% • 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7131 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 457aa long hypothetical protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	435	3341	2143	553	627	3	15	0	0	0
1	B	434	3330	2135	552	625	3	15	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	122	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	133	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	185	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	194	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	220	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	279	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	282	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	299	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	338	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	347	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	371	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	384	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	409	MSE	MET	MODIFIED RESIDUE	UNP O58493
A	411	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	1	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	122	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	133	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	185	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	194	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	220	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	279	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	282	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	299	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	338	MSE	MET	MODIFIED RESIDUE	UNP O58493

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	347	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	371	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	384	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	409	MSE	MET	MODIFIED RESIDUE	UNP O58493
B	411	MSE	MET	MODIFIED RESIDUE	UNP O58493

- Molecule 2 is water.

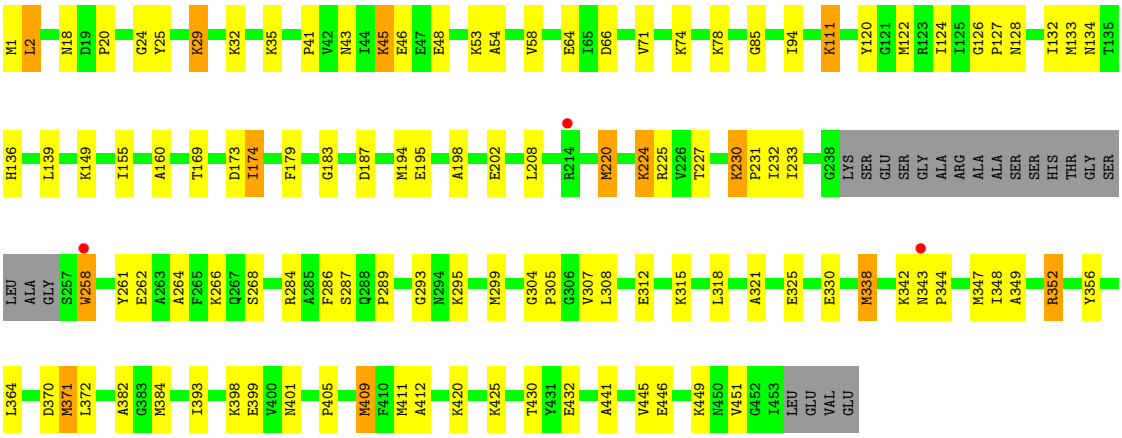
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	287	287	287	0	0
			Total	O		
2	B	173	173	173	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

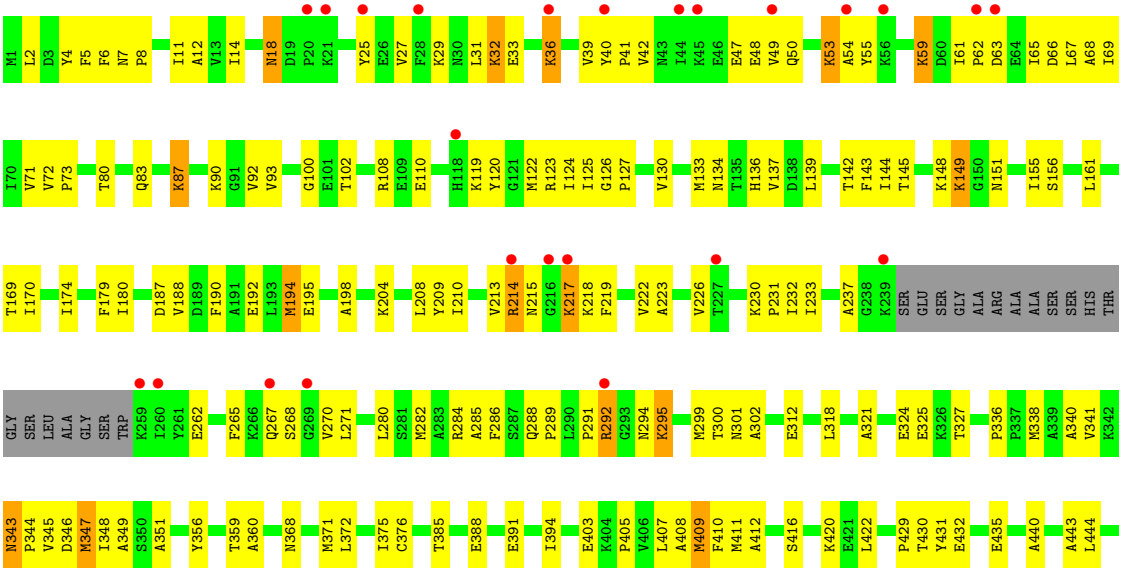
- Molecule 1: 457aa long hypothetical protein

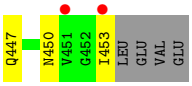
Chain A:



- Molecule 1: 457aa long hypothetical protein

Chain B:




Q447
N450
V451
G452
I453
LEU
GLU
VAL
GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	143.46Å 70.59Å 88.51Å 90.00° 90.66° 90.00°	Depositor
Resolution (Å)	19.85 – 2.20 19.85 – 2.19	Depositor EDS
% Data completeness (in resolution range)	99.9 (19.85-2.20) 96.1 (19.85-2.19)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.71 (at 2.19Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.246 , 0.261 0.246 , 0.261	Depositor DCC
R_{free} test set	2251 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	35.4	Xtriage
Anisotropy	0.497	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 31.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.135 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7131	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.16% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.64	3/3379 (0.1%)	0.82	5/4528 (0.1%)
1	B	0.59	2/3366 (0.1%)	0.70	0/4508
All	All	0.62	5/6745 (0.1%)	0.76	5/9036 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	258	TRP	NE1-CE2	8.74	1.49	1.37
1	A	194	MSE	SE-CE	-6.31	1.58	1.95
1	A	371	MSE	SE-CE	-5.85	1.60	1.95
1	B	194	MSE	CG-SE	-5.21	1.77	1.95
1	B	371	MSE	CG-SE	-5.05	1.78	1.95

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	330	GLU	CA-CB-CG	6.07	126.75	113.40
1	A	330	GLU	CG-CD-OE1	5.87	130.03	118.30
1	A	330	GLU	CG-CD-OE2	-5.71	106.88	118.30
1	A	225	ARG	CD-NE-CZ	-5.20	116.32	123.60
1	A	352	ARG	CD-NE-CZ	-5.04	116.55	123.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	214	ARG	Sidechain

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	3341	0	3460	102	0
1	B	3330	0	3458	181	0
2	A	287	0	0	6	0
2	B	173	0	0	8	0
All	All	7131	0	6918	278	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 20.

The worst 5 of 278 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:409:MSE:HE1	1:A:411:MSE:SE	2.16	0.95
1:B:156:SER:HB2	1:B:209:TYR:HB3	1.48	0.95
1:B:411:MSE:HB3	1:B:412:ALA:HB2	1.49	0.95
1:A:343:ASN:HB2	1:A:344:PRO:HD3	1.58	0.86
1:B:136:HIS:HA	1:B:149:LYS:HE2	1.59	0.85

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	431/457 (94%)	410 (95%)	21 (5%)	0	100	100
1	B	430/457 (94%)	394 (92%)	34 (8%)	2 (0%)	29	31
All	All	861/914 (94%)	804 (93%)	55 (6%)	2 (0%)	47	55

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	25	TYR
1	B	262	GLU

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	354/354 (100%)	336 (95%)	18 (5%)	24	29
1	B	353/354 (100%)	335 (95%)	18 (5%)	24	29
All	All	707/708 (100%)	671 (95%)	36 (5%)	24	29

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	267	GLN
1	B	409	MSE
1	B	292	ARG
1	B	343	ASN
1	A	261	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	18	ASN
1	B	50	GLN
1	B	365	GLN
1	B	343	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	401	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 monosaccharides 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, 95th 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^2)	Q<0.9
1	A	420/457 (91%)	-0.22	3 (0%)	87	86	19, 33, 50, 68	0
1	B	419/457 (91%)	0.43	26 (6%)	20	19	28, 53, 83, 88	0
All	All	839/914 (91%)	0.11	29 (3%)	44	42	19, 42, 74, 88	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	63	ASP	5.5
1	B	62	PRO	5.3
1	B	49	VAL	4.9
1	B	56	LYS	4.5
1	B	453	ILE	4.4

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.