wwPDB X-ray Structure Validation Summary Report (i)

Nov 16, 2023 – 02:20 AM JST

PDB ID	:	6KJG
Title	:	Crystal structure of PsoF
Authors	:	Hara, K.; Hashimoto, H.; Matsushita, T.; Tsunematsu, Y.; Watanabe, K.
Deposited on	:	2019-07-22
Resolution	:	1.99 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 (1)) were used in the production of this report:

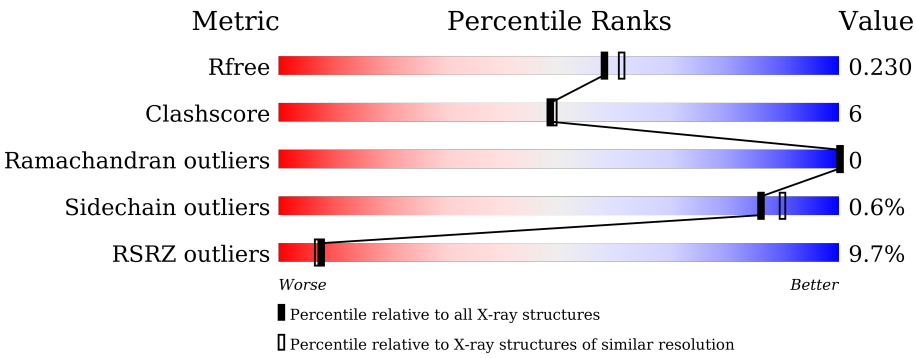
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.99 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	378	10% 80% 12% • 8%
1	B	378	9% 80% 12% 8%
1	C	378	8% 80% 11% • 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8613 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 Dual-functional monooxygenase/methyltransferase psoF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	348	2707	1709	471	515	12	0	0	0
			Total	C	N	O	S			
1	B	349	2713	1712	472	517	12	0	0	0
			Total	C	N	O	S			
1	C	348	2706	1708	471	515	12	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	537	MET	-	expression tag	UNP Q4WAZ0
A	906	GLY	-	expression tag	UNP Q4WAZ0
A	907	HIS	-	expression tag	UNP Q4WAZ0
A	908	HIS	-	expression tag	UNP Q4WAZ0
A	909	HIS	-	expression tag	UNP Q4WAZ0
A	910	HIS	-	expression tag	UNP Q4WAZ0
A	911	HIS	-	expression tag	UNP Q4WAZ0
A	912	HIS	-	expression tag	UNP Q4WAZ0
A	913	HIS	-	expression tag	UNP Q4WAZ0
A	914	HIS	-	expression tag	UNP Q4WAZ0
B	537	MET	-	expression tag	UNP Q4WAZ0
B	906	GLY	-	expression tag	UNP Q4WAZ0
B	907	HIS	-	expression tag	UNP Q4WAZ0
B	908	HIS	-	expression tag	UNP Q4WAZ0
B	909	HIS	-	expression tag	UNP Q4WAZ0
B	910	HIS	-	expression tag	UNP Q4WAZ0
B	911	HIS	-	expression tag	UNP Q4WAZ0
B	912	HIS	-	expression tag	UNP Q4WAZ0
B	913	HIS	-	expression tag	UNP Q4WAZ0
B	914	HIS	-	expression tag	UNP Q4WAZ0
C	537	MET	-	expression tag	UNP Q4WAZ0
C	906	GLY	-	expression tag	UNP Q4WAZ0
C	907	HIS	-	expression tag	UNP Q4WAZ0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	908	HIS	-	expression tag	UNP Q4WAZ0
C	909	HIS	-	expression tag	UNP Q4WAZ0
C	910	HIS	-	expression tag	UNP Q4WAZ0
C	911	HIS	-	expression tag	UNP Q4WAZ0
C	912	HIS	-	expression tag	UNP Q4WAZ0
C	913	HIS	-	expression tag	UNP Q4WAZ0
C	914	HIS	-	expression tag	UNP Q4WAZ0

- Molecule 2 is water.

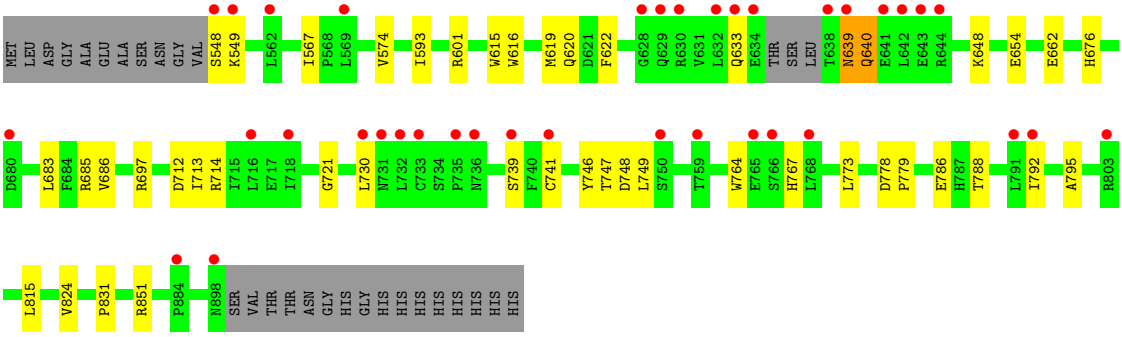
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	156	156	156	0	0
			Total	O		
2	B	149	149	149	0	0
			Total	O		
2	C	182	182	182	0	0

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.

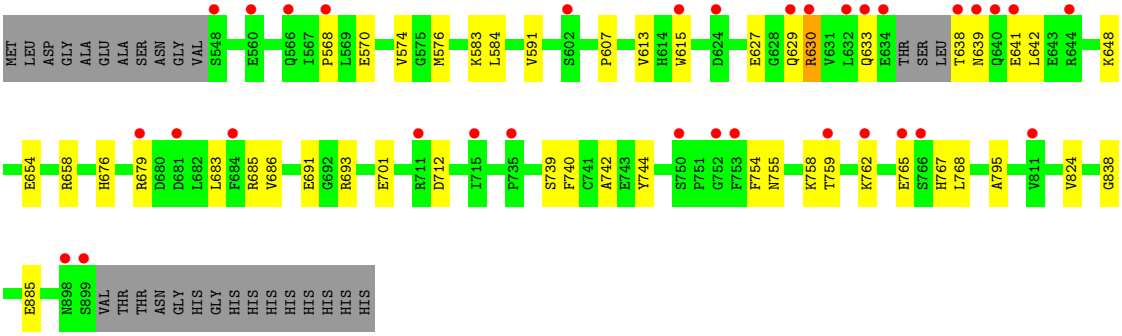
• Molecule 1: Dual-functional monooxygenase/methyltransferase psoF

Chain A:



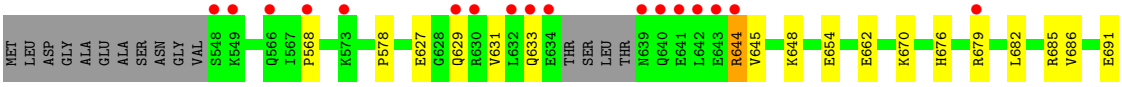
• Molecule 1: Dual-functional monooxygenase/methyltransferase psoF

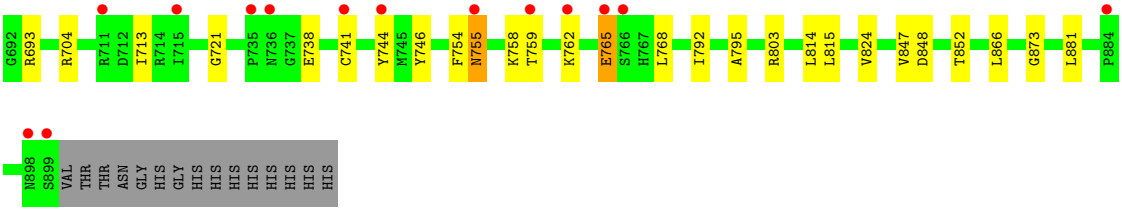
Chain B:



• Molecule 1: Dual-functional monooxygenase/methyltransferase psoF

Chain C:




G692
R693
R704
R711
D712
I713
R714
I715
G721
P735
N736
G737
E738
C741
Y744
M745
Y746
F754
N755
K758
T759
K762
E765
S766
H767
L768
I792
A795
R803
L814
L815
V824
V847
D848
T852
L866
G873
L881
P884
N898
S899
VAL
THR
THR
ASN
GLY
HIS
GLY
HIS
HIS
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	189.24Å 44.10Å 142.62Å 90.00° 93.59° 90.00°	Depositor
Resolution (Å)	19.80 – 1.99 19.81 – 1.99	Depositor EDS
% Data completeness (in resolution range)	98.5 (19.80-1.99) 98.5 (19.81-1.99)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.95 (at 1.99Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.197 , 0.230 0.198 , 0.230	Depositor DCC
R_{free} test set	4037 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	33.1	Xtriage
Anisotropy	0.421	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 50.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8613	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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	0.40	0/2761	0.58	0/3743
1	B	0.38	0/2767	0.55	0/3751
1	C	0.41	0/2760	0.58	0/3741
All	All	0.40	0/8288	0.57	0/11235

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	A	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	639	ASN	Peptide
1	A	640	GLN	Peptide

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	2707	0	2685	33	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2713	0	2690	33	0
1	C	2706	0	2683	31	0
2	A	156	0	0	3	1
2	B	149	0	0	6	0
2	C	182	0	0	8	1
All	All	8613	0	8058	97	2

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 97 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:C:803:ARG:NH1	2:C:1002:HOH:O	2.18	0.75
1:A:549:LYS:N	1:A:549:LYS:HD3	2.05	0.70
1:C:765:GLU:O	1:C:765:GLU:HG3	1.91	0.69
1:C:644:ARG:CZ	1:C:648:LYS:HE3	2.23	0.68
1:B:740:PHE:H	1:B:767:HIS:HE1	1.42	0.66

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:1065:HOH:O	2:A:1083:HOH:O[2_555]	2.05	0.15
2:C:1057:HOH:O	2:C:1137:HOH:O[2_556]	2.11	0.09

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	344/378 (91%)	335 (97%)	9 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	345/378 (91%)	337 (98%)	8 (2%)	0	100	100
1	C	344/378 (91%)	336 (98%)	8 (2%)	0	100	100
All	All	1033/1134 (91%)	1008 (98%)	25 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	290/314 (92%)	290 (100%)	0	100	100
1	B	291/314 (93%)	289 (99%)	2 (1%)	84	88
1	C	290/314 (92%)	287 (99%)	3 (1%)	76	81
All	All	871/942 (92%)	866 (99%)	5 (1%)	86	90

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

Mol	Chain	Res	Type
1	B	630	ARG
1	B	765	GLU
1	C	644	ARG
1	C	755	ASN
1	C	765	GLU

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

Mol	Chain	Res	Type
1	B	767	HIS
1	C	633	GLN

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	348/378 (92%)	0.66	37 (10%)	6	5	21, 39, 63, 81	0
1	B	349/378 (92%)	0.62	33 (9%)	8	7	23, 40, 63, 89	0
1	C	348/378 (92%)	0.57	31 (8%)	9	8	19, 36, 63, 87	0
All	All	1045/1134 (92%)	0.62	101 (9%)	7	7	19, 38, 63, 89	0

The worst 5 of 101 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	638	THR	10.0
1	A	638	THR	7.1
1	C	639	ASN	6.9
1	C	633	GLN	6.5
1	C	640	GLN	6.2

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.