



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 01:56 am GMT

PDB ID : 4W8Y
Title : Structure of full length Cmr2 from *Pyrococcus furiosus* (Manganese bound form)
Authors : Benda, C.; Ebert, J.; Baumgaertner, M.; Conti, E.
Deposited on : 2014-08-26
Resolution : 3.00 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

<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

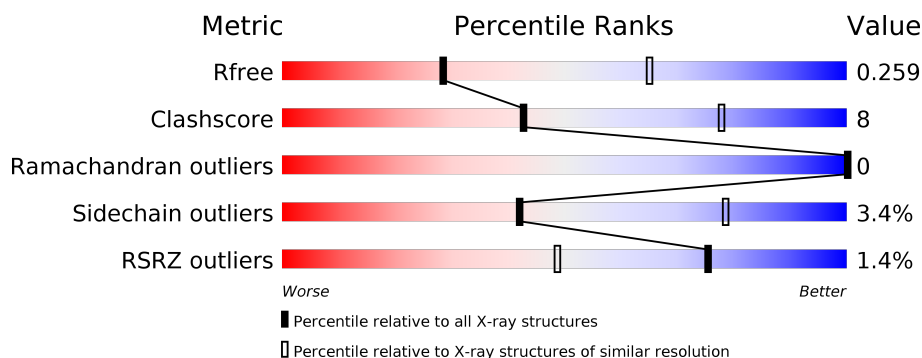
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 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	888	<div> <div> <div></div> <div>75%</div> <div>18%</div> <div>• 6%</div> </div> <div>2%</div> </div>
1	B	888	<div> <div> <div></div> <div>73%</div> <div>20%</div> <div>• 7%</div> </div> <div>2%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ZN	B	901	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13055 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 CRISPR system Cmr subunit Cmr2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	835	Total	C	N	O	S	0	0	0
			6498	4211	1076	1198	13			
1	B	828	Total	C	N	O	S	0	0	0
			6517	4235	1069	1199	14			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-16	MET	-	initiating methionine	UNP Q8U1S6
A	-15	HIS	-	expression tag	UNP Q8U1S6
A	-14	HIS	-	expression tag	UNP Q8U1S6
A	-13	HIS	-	expression tag	UNP Q8U1S6
A	-12	HIS	-	expression tag	UNP Q8U1S6
A	-11	HIS	-	expression tag	UNP Q8U1S6
A	-10	HIS	-	expression tag	UNP Q8U1S6
A	-9	GLU	-	expression tag	UNP Q8U1S6
A	-8	ASN	-	expression tag	UNP Q8U1S6
A	-7	LEU	-	expression tag	UNP Q8U1S6
A	-6	TYR	-	expression tag	UNP Q8U1S6
A	-5	PHE	-	expression tag	UNP Q8U1S6
A	-4	GLN	-	expression tag	UNP Q8U1S6
A	-3	GLY	-	expression tag	UNP Q8U1S6
A	-2	ALA	-	expression tag	UNP Q8U1S6
A	-1	ALA	-	expression tag	UNP Q8U1S6
A	0	SER	-	expression tag	UNP Q8U1S6
B	-16	MET	-	initiating methionine	UNP Q8U1S6
B	-15	HIS	-	expression tag	UNP Q8U1S6
B	-14	HIS	-	expression tag	UNP Q8U1S6
B	-13	HIS	-	expression tag	UNP Q8U1S6
B	-12	HIS	-	expression tag	UNP Q8U1S6
B	-11	HIS	-	expression tag	UNP Q8U1S6
B	-10	HIS	-	expression tag	UNP Q8U1S6
B	-9	GLU	-	expression tag	UNP Q8U1S6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-8	ASN	-	expression tag	UNP Q8U1S6
B	-7	LEU	-	expression tag	UNP Q8U1S6
B	-6	TYR	-	expression tag	UNP Q8U1S6
B	-5	PHE	-	expression tag	UNP Q8U1S6
B	-4	GLN	-	expression tag	UNP Q8U1S6
B	-3	GLY	-	expression tag	UNP Q8U1S6
B	-2	ALA	-	expression tag	UNP Q8U1S6
B	-1	ALA	-	expression tag	UNP Q8U1S6
B	0	SER	-	expression tag	UNP Q8U1S6

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Zn	0	0
			1	1		
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	4	Total	Mn	0	0
			4	4		
3	A	4	Total	Mn	0	0
			4	4		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	16	Total	O	0	0
			16	16		
4	B	14	Total	O	0	0
			14	14		

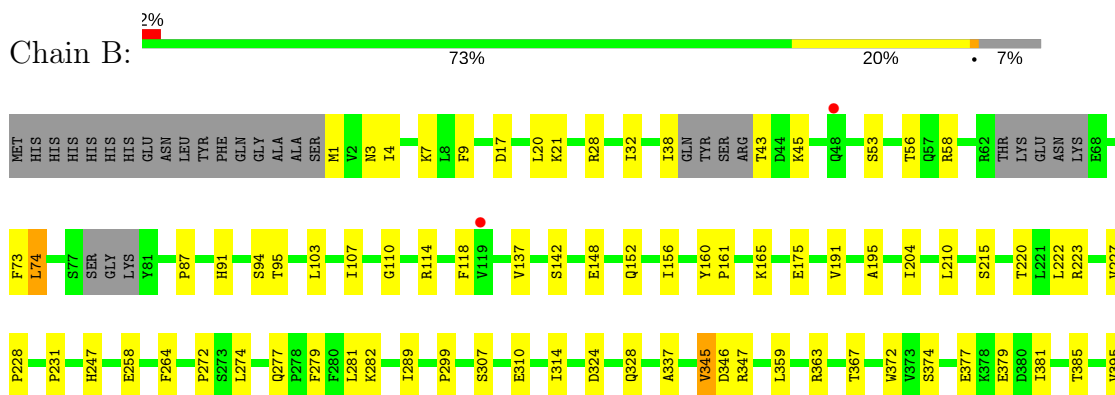
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: CRISPR system Cmr subunit Cmr2



• Molecule 1: CRISPR system Cmr subunit Cmr2



F791	L792	Y793	H794	V795	E798	V799	D800	T801	W802	P803	LYS	VAL	GLY	T807	F813	F814	V815	I816	I819	ARG	GLY	ARG	ASN	R824	E825	E826	T827	L830	R831	L842	V845	E850	K853	V854	L857	F860	L861	K862	K863	I864	P871					
L649	S653	E656	S659	V660	D663	E664	Y669	A670	G671	D672	D673	D674	V675	L676	A677	I678	D682	E694	S698	I714	Y719	L727	L734	N735	K739	T746	L747	V763	E766	L767	L768	R769	E782	GLU	LYS	GLY	GLY	VAL	GLY	LYS	R790					
Y522	L543	L548	Y549	S556	GLY	LEU	SER	LYS	LYS	LEU	LYS	ASN	LYS	GLU	ILE	D569	V573	V577	D578	F579	L580	N581	Y584	K592	I596	L597	V598	G614	E615	I616	S617	T618	R619	I620	H621	P622	R625	I630	P631	K634	Y635	Y636	S637	T638	P639	Q640
L399	K400	F401	A402	L411	Y417	P418	L424	R430	K431	V432	T433	E434	E435	ARG	PHE	GLU	LYS	SER	E441	C448	H449	V450	C451	I457	F458	G459	D460	L467	W471	P476	L477	C478	L482	R492	K498	L499	R500	V504	V507	A508	L509	R515				

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	62.40Å 167.93Å 100.91Å 90.00° 98.91° 90.00°	Depositor
Resolution (Å)	85.72 – 3.00 61.65 – 3.00	Depositor EDS
% Data completeness (in resolution range)	98.4 (85.72-3.00) 98.5 (61.65-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.36 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1678)	Depositor
R, R_{free}	0.223 , 0.259 0.225 , 0.259	Depositor DCC
R_{free} test set	3949 reflections (4.99%)	DCC
Wilson B-factor (Å ²)	90.0	Xtriage
Anisotropy	0.282	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 57.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	13055	wwPDB-VP
Average B, all atoms (Å ²)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.05% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

5.1 Standard geometry

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

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	0.24	0/6634	0.42	0/9002
1	B	0.24	0/6656	0.41	0/9024
All	All	0.24	0/13290	0.41	0/18026

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

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	6498	0	6297	94	0
1	B	6517	0	6365	107	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	4	0	0	0	0
3	B	4	0	0	0	0
4	A	16	0	0	1	0
4	B	14	0	0	2	0
All	All	13055	0	12662	201	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 8.

The worst 5 of 201 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:360:LYS:NZ	1:A:440:SER:OG	2.07	0.86
1:A:210:LEU:HD11	1:A:639:PRO:HB2	1.63	0.80
1:B:21:LYS:O	1:B:28:ARG:NH2	2.19	0.73
1:A:21:LYS:O	1:A:28:ARG:NH2	2.21	0.72
1:B:656:GLU:HB3	1:B:694:GLU:HG2	1.71	0.71

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	819/888 (92%)	797 (97%)	22 (3%)	0	100	100
1	B	810/888 (91%)	789 (97%)	21 (3%)	0	100	100
All	All	1629/1776 (92%)	1586 (97%)	43 (3%)	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	666/795 (84%)	641 (96%)	25 (4%)	38	75

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	677/795 (85%)	656 (97%)	21 (3%)	45 80
All	All	1343/1590 (84%)	1297 (97%)	46 (3%)	42 78

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	796	LEU
1	B	1	MET
1	B	746	THR
1	A	800	ASP
1	A	810	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	152	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 10 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

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

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, 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	835/888 (94%)	-0.14	9 (1%)	80 55	53, 89, 134, 213	0
1	B	828/888 (93%)	-0.11	14 (1%)	70 42	58, 92, 129, 192	0
All	All	1663/1776 (93%)	-0.12	23 (1%)	75 49	53, 91, 131, 213	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	435	GLU	13.2
1	B	432	VAL	6.6
1	B	434	GLU	6.6
1	B	431	LYS	5.7
1	A	627	TYR	5.1

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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(\AA^2)	Q<0.9
2	ZN	B	901	1/1	0.96	0.19	2.84	85,85,85,85	0
3	MN	B	902	1/1	0.91	0.17	-0.63	143,143,143,143	0
2	ZN	A	901	1/1	0.99	0.15	-0.99	64,64,64,64	0
3	MN	A	905	1/1	0.58	0.13	-	132,132,132,132	0
3	MN	A	904	1/1	0.96	0.19	-	76,76,76,76	0
3	MN	B	905	1/1	0.81	0.23	-	191,191,191,191	0
3	MN	A	902	1/1	0.81	0.23	-	172,172,172,172	0
3	MN	A	903	1/1	0.93	0.18	-	153,153,153,153	0
3	MN	B	903	1/1	0.24	0.14	-	168,168,168,168	0
3	MN	B	904	1/1	0.98	0.23	-	82,82,82,82	0

6.5 Other polymers [i](#)

There are no such residues in this entry.