



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 06:28 am GMT

PDB ID : 2F4P
Title : Crystal structure of a cupin-like protein (tm1010) from thermotoga maritima
msb8 at 1.90 Å resolution
Authors : Joint Center for Structural Genomics (JCSG)
Deposited on : 2005-11-23
Resolution : 1.90 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

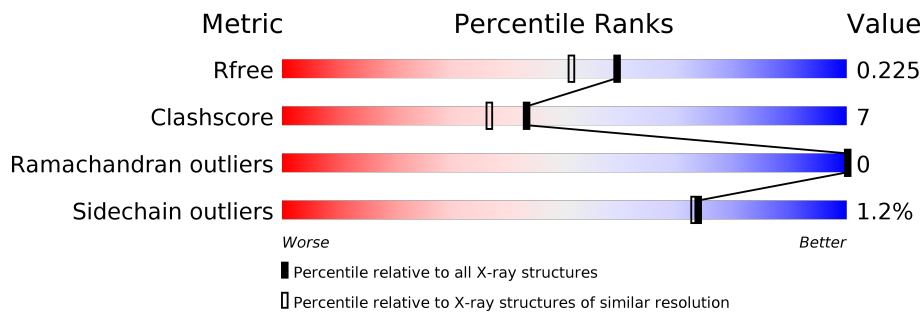
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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}	111664	5502 (1.90-1.90)
Clashscore	122126	6115 (1.90-1.90)
Ramachandran outliers	120053	6048 (1.90-1.90)
Sidechain outliers	120020	6048 (1.90-1.90)

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	147	
1	B	147	
1	C	147	
1	D	147	

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
3	EDO	B	139	-	-	X	-
3	EDO	C	138	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4826 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 TM1010.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	134	Total	C	N	O	Se	0	4	0
			1065	688	183	193	1			
1	B	132	Total	C	N	O	Se	0	6	0
			1064	683	188	192	1			
1	C	133	Total	C	N	O	Se	0	5	0
			1056	680	185	190	1			
1	D	133	Total	C	N	O	Se	0	6	0
			1058	684	180	193	1			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	LEADER SEQUENCE	GB 4981537
A	-10	GLY	-	LEADER SEQUENCE	GB 4981537
A	-9	SER	-	LEADER SEQUENCE	GB 4981537
A	-8	ASP	-	LEADER SEQUENCE	GB 4981537
A	-7	LYS	-	LEADER SEQUENCE	GB 4981537
A	-6	ILE	-	LEADER SEQUENCE	GB 4981537
A	-5	HIS	-	LEADER SEQUENCE	GB 4981537
A	-4	HIS	-	LEADER SEQUENCE	GB 4981537
A	-3	HIS	-	LEADER SEQUENCE	GB 4981537
A	-2	HIS	-	LEADER SEQUENCE	GB 4981537
A	-1	HIS	-	LEADER SEQUENCE	GB 4981537
A	0	HIS	-	LEADER SEQUENCE	GB 4981537
A	1	MSE	MET	MODIFIED RESIDUE	GB 4981537
A	25	MSE	MET	MODIFIED RESIDUE	GB 4981537
B	-11	MSE	-	LEADER SEQUENCE	GB 4981537
B	-10	GLY	-	LEADER SEQUENCE	GB 4981537
B	-9	SER	-	LEADER SEQUENCE	GB 4981537
B	-8	ASP	-	LEADER SEQUENCE	GB 4981537
B	-7	LYS	-	LEADER SEQUENCE	GB 4981537
B	-6	ILE	-	LEADER SEQUENCE	GB 4981537
B	-5	HIS	-	LEADER SEQUENCE	GB 4981537

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	HIS	-	LEADER SEQUENCE	GB 4981537
B	-3	HIS	-	LEADER SEQUENCE	GB 4981537
B	-2	HIS	-	LEADER SEQUENCE	GB 4981537
B	-1	HIS	-	LEADER SEQUENCE	GB 4981537
B	0	HIS	-	LEADER SEQUENCE	GB 4981537
B	1	MSE	MET	MODIFIED RESIDUE	GB 4981537
B	25	MSE	MET	MODIFIED RESIDUE	GB 4981537
C	-11	MSE	-	LEADER SEQUENCE	GB 4981537
C	-10	GLY	-	LEADER SEQUENCE	GB 4981537
C	-9	SER	-	LEADER SEQUENCE	GB 4981537
C	-8	ASP	-	LEADER SEQUENCE	GB 4981537
C	-7	LYS	-	LEADER SEQUENCE	GB 4981537
C	-6	ILE	-	LEADER SEQUENCE	GB 4981537
C	-5	HIS	-	LEADER SEQUENCE	GB 4981537
C	-4	HIS	-	LEADER SEQUENCE	GB 4981537
C	-3	HIS	-	LEADER SEQUENCE	GB 4981537
C	-2	HIS	-	LEADER SEQUENCE	GB 4981537
C	-1	HIS	-	LEADER SEQUENCE	GB 4981537
C	0	HIS	-	LEADER SEQUENCE	GB 4981537
C	1	MSE	MET	MODIFIED RESIDUE	GB 4981537
C	25	MSE	MET	MODIFIED RESIDUE	GB 4981537
D	-11	MSE	-	LEADER SEQUENCE	GB 4981537
D	-10	GLY	-	LEADER SEQUENCE	GB 4981537
D	-9	SER	-	LEADER SEQUENCE	GB 4981537
D	-8	ASP	-	LEADER SEQUENCE	GB 4981537
D	-7	LYS	-	LEADER SEQUENCE	GB 4981537
D	-6	ILE	-	LEADER SEQUENCE	GB 4981537
D	-5	HIS	-	LEADER SEQUENCE	GB 4981537
D	-4	HIS	-	LEADER SEQUENCE	GB 4981537
D	-3	HIS	-	LEADER SEQUENCE	GB 4981537
D	-2	HIS	-	LEADER SEQUENCE	GB 4981537
D	-1	HIS	-	LEADER SEQUENCE	GB 4981537
D	0	HIS	-	LEADER SEQUENCE	GB 4981537
D	1	MSE	MET	MODIFIED RESIDUE	GB 4981537
D	25	MSE	MET	MODIFIED RESIDUE	GB 4981537

- Molecule 2 is UNKNOWN LIGAND (three-letter code: UNL) (formula:).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total O 9 9	0	0
2	A	1	Total O 10 10	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total O 9 9	0	0
2	C	1	Total O 9 9	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



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

Continued on next page...

Continued from previous page...

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

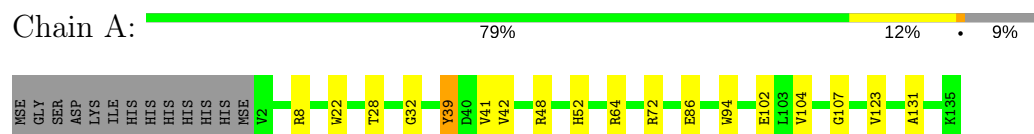
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	137	Total	O	0	1
			138	138		
4	B	117	Total	O	0	1
			118	118		
4	C	119	Total	O	0	1
			120	120		
4	D	129	Total	O	0	1
			130	130		

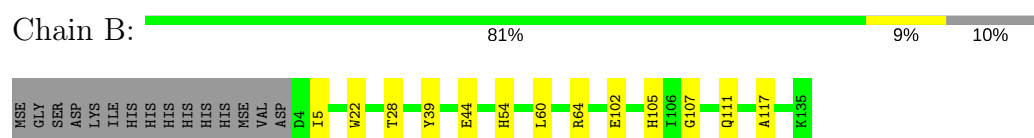
3 Residue-property plots [i](#)

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. 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. 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.

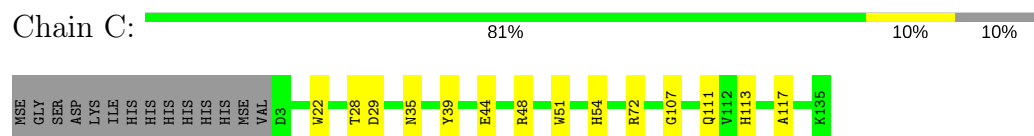
- Molecule 1: hypothetical protein TM1010



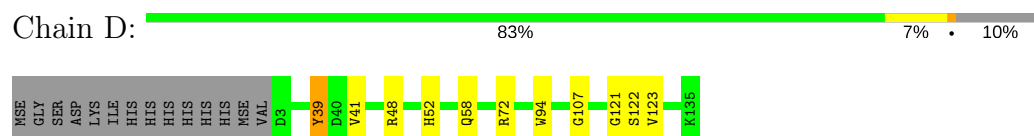
- Molecule 1: hypothetical protein TM1010



- Molecule 1: hypothetical protein TM1010



- Molecule 1: hypothetical protein TM1010



4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	71.75Å 71.75Å 188.60Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	21.54 – 1.90 21.54 – 1.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (21.54-1.90) 100.0 (21.54-1.90)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.70 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.160 , 0.216 0.174 , 0.225	Depositor DCC
R_{free} test set	2172 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	23.8	Xtriage
Anisotropy	0.141	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 34.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.52$, $\langle L^2 \rangle = 0.36$	Xtriage
Estimated twinning fraction	0.478 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4826	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 5.27% 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: UNL, EDO

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	0.62	0/1109	0.71	2/1510 (0.1%)
1	B	0.54	0/1105	0.65	0/1502
1	C	0.54	0/1096	0.62	0/1489
1	D	0.60	0/1107	0.69	1/1506 (0.1%)
All	All	0.57	0/4417	0.67	3/6007 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	72	ARG	NE-CZ-NH2	-7.69	116.45	120.30
1	A	72	ARG	NE-CZ-NH1	7.49	124.05	120.30
1	D	72	ARG	NE-CZ-NH2	-5.11	117.75	120.30

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	1065	0	1011	18	0
1	B	1064	0	1011	13	0
1	C	1056	0	1009	17	0
1	D	1058	0	1000	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	10	0	0	1	0
2	B	9	0	0	0	0
2	C	9	0	0	0	0
2	D	9	0	0	1	0
3	A	8	0	12	3	0
3	B	16	0	24	8	0
3	C	8	0	12	9	0
3	D	8	0	12	2	0
4	A	138	0	0	2	0
4	B	118	0	0	5	0
4	C	120	0	0	3	0
4	D	130	0	0	0	0
All	All	4826	0	4091	61	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 7.

The worst 5 of 61 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:42[B]:VAL:HG12	1:A:104:VAL:HG22	1.61	0.82
1:A:39[B]:TYR:HD1	1:A:41[B]:VAL:HG13	1.47	0.77
1:A:39[B]:TYR:CD1	1:A:41[B]:VAL:HG13	2.22	0.75
1:C:29:ASP:H	3:C:138:EDO:H21	1.60	0.66
1:D:39[B]:TYR:HD1	1:D:41[B]:VAL:HG13	1.59	0.66

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	136/147 (92%)	135 (99%)	1 (1%)	0	100	100
1	B	136/147 (92%)	136 (100%)	0	0	100	100
1	C	136/147 (92%)	136 (100%)	0	0	100	100
1	D	137/147 (93%)	136 (99%)	1 (1%)	0	100	100
All	All	545/588 (93%)	543 (100%)	2 (0%)	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	109/121 (90%)	106 (97%)	3 (3%)	47	39
1	B	109/121 (90%)	109 (100%)	0	100	100
1	C	107/121 (88%)	106 (99%)	1 (1%)	81	81
1	D	108/121 (89%)	105 (97%)	3 (3%)	47	39
All	All	433/484 (90%)	426 (98%)	7 (2%)	74	62

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	48	ARG
1	D	48	ARG
1	D	39[A]	TYR
1	A	39[B]	TYR
1	D	39[B]	TYR

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

Mol	Chain	Res	Type
1	A	58	GLN
1	D	58	GLN

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 ⓘ

Of 14 ligands modelled in this entry, 4 are unknown - leaving 10 for Mogul analysis.

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	$\# Z > 2$
3	EDO	A	137	-	3,3,3	0.41	0	2,2,2	0.76	0
3	EDO	A	138	-	3,3,3	0.32	0	2,2,2	0.62	0
3	EDO	B	137	-	3,3,3	0.51	0	2,2,2	0.32	0
3	EDO	B	138	-	3,3,3	0.66	0	2,2,2	0.11	0
3	EDO	B	139	-	3,3,3	0.43	0	2,2,2	0.22	0
3	EDO	B	140	-	3,3,3	0.53	0	2,2,2	0.48	0
3	EDO	C	137	-	3,3,3	0.46	0	2,2,2	0.45	0
3	EDO	C	138	-	3,3,3	0.52	0	2,2,2	0.66	0
3	EDO	D	137	-	3,3,3	0.28	0	2,2,2	0.61	0
3	EDO	D	138	-	3,3,3	0.47	0	2,2,2	0.57	0

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
3	EDO	A	137	-	-	0/1/1/1	0/0/0/0
3	EDO	A	138	-	-	0/1/1/1	0/0/0/0
3	EDO	B	137	-	-	0/1/1/1	0/0/0/0
3	EDO	B	138	-	-	0/1/1/1	0/0/0/0
3	EDO	B	139	-	-	0/1/1/1	0/0/0/0
3	EDO	B	140	-	-	0/1/1/1	0/0/0/0
3	EDO	C	137	-	-	0/1/1/1	0/0/0/0
3	EDO	C	138	-	-	0/1/1/1	0/0/0/0
3	EDO	D	137	-	-	0/1/1/1	0/0/0/0
3	EDO	D	138	-	-	0/1/1/1	0/0/0/0

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.

5 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	138	EDO	3	0
3	B	139	EDO	5	0
3	B	140	EDO	3	0
3	C	138	EDO	9	0
3	D	138	EDO	2	0

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.