wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 09:05 pm BST

PDB ID	:	1YBX
Title	:	Conserved hypothetical protein Cth-383 from Clostridium thermocellum
Authors	:	Tempel, W.; Chang, J.; Zhao, M.; Habel, J.; Kataeva, I.; Xu, H.; Chen, L.; Lee, D.; Nguyen, J.; Chang, S.-H.; Horanyi, P.; Florence, Q.; Zhou, W.; Lin, D.; Zhang, H.; Ljundahl, L.; Liu, Z.-J.; Rose, J.; Wang, B.-C.; Southeast Collaboratory for Structural Genomics (SECSG)
Deposited on	:	2004-12-21
Resolution	:	1.80 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 following versions of software and data (see references (1)) were used in the production of this report:

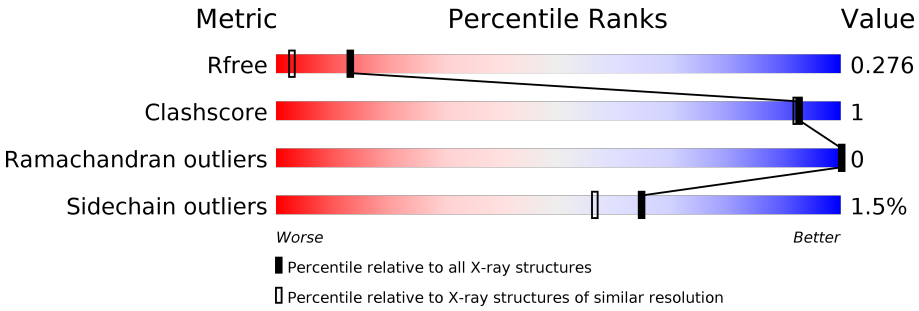
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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 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%

Mol	Chain	Length	Quality of chain
1	A	143	59% • • 36%
1	B	143	62% • 36%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 1439 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 Conserved hypothetical protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	91	Total	C	N	O	Se	0	0	0
			673	418	112	139	4			
1	B	92	Total	C	N	O	Se	0	1	0
			672	419	113	135	5			

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-29	MSE	-	CLONING ARTIFACT	GB 48857448
A	-28	GLY	-	CLONING ARTIFACT	GB 48857448
A	-27	SER	-	CLONING ARTIFACT	GB 48857448
A	-26	SER	-	CLONING ARTIFACT	GB 48857448
A	-25	HIS	-	CLONING ARTIFACT	GB 48857448
A	-24	HIS	-	CLONING ARTIFACT	GB 48857448
A	-23	HIS	-	CLONING ARTIFACT	GB 48857448
A	-22	HIS	-	CLONING ARTIFACT	GB 48857448
A	-21	HIS	-	CLONING ARTIFACT	GB 48857448
A	-20	HIS	-	CLONING ARTIFACT	GB 48857448
A	-19	SER	-	CLONING ARTIFACT	GB 48857448
A	-18	SER	-	CLONING ARTIFACT	GB 48857448
A	-17	GLY	-	CLONING ARTIFACT	GB 48857448
A	-16	LEU	-	CLONING ARTIFACT	GB 48857448
A	-15	VAL	-	CLONING ARTIFACT	GB 48857448
A	-14	PRO	-	CLONING ARTIFACT	GB 48857448
A	-13	ARG	-	CLONING ARTIFACT	GB 48857448
A	-12	GLY	-	CLONING ARTIFACT	GB 48857448
A	-11	SER	-	CLONING ARTIFACT	GB 48857448
A	-10	GLN	-	CLONING ARTIFACT	GB 48857448
A	-9	SER	-	CLONING ARTIFACT	GB 48857448
A	-8	THR	-	CLONING ARTIFACT	GB 48857448
A	-7	SER	-	CLONING ARTIFACT	GB 48857448
A	-6	LEU	-	CLONING ARTIFACT	GB 48857448
A	-5	TYR	-	CLONING ARTIFACT	GB 48857448

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	LYS	-	CLONING ARTIFACT	GB 48857448
A	-3	LYS	-	CLONING ARTIFACT	GB 48857448
A	-2	ALA	-	CLONING ARTIFACT	GB 48857448
A	-1	GLY	-	CLONING ARTIFACT	GB 48857448
A	0	LEU	-	CLONING ARTIFACT	GB 48857448
A	1	MSE	MET	MODIFIED RESIDUE	GB 48857448
A	23	MSE	MET	MODIFIED RESIDUE	GB 48857448
A	27	MSE	MET	MODIFIED RESIDUE	GB 48857448
A	75	MSE	MET	MODIFIED RESIDUE	GB 48857448
A	94	MSE	MET	MODIFIED RESIDUE	GB 48857448
B	-29	MSE	-	CLONING ARTIFACT	GB 48857448
B	-28	GLY	-	CLONING ARTIFACT	GB 48857448
B	-27	SER	-	CLONING ARTIFACT	GB 48857448
B	-26	SER	-	CLONING ARTIFACT	GB 48857448
B	-25	HIS	-	CLONING ARTIFACT	GB 48857448
B	-24	HIS	-	CLONING ARTIFACT	GB 48857448
B	-23	HIS	-	CLONING ARTIFACT	GB 48857448
B	-22	HIS	-	CLONING ARTIFACT	GB 48857448
B	-21	HIS	-	CLONING ARTIFACT	GB 48857448
B	-20	HIS	-	CLONING ARTIFACT	GB 48857448
B	-19	SER	-	CLONING ARTIFACT	GB 48857448
B	-18	SER	-	CLONING ARTIFACT	GB 48857448
B	-17	GLY	-	CLONING ARTIFACT	GB 48857448
B	-16	LEU	-	CLONING ARTIFACT	GB 48857448
B	-15	VAL	-	CLONING ARTIFACT	GB 48857448
B	-14	PRO	-	CLONING ARTIFACT	GB 48857448
B	-13	ARG	-	CLONING ARTIFACT	GB 48857448
B	-12	GLY	-	CLONING ARTIFACT	GB 48857448
B	-11	SER	-	CLONING ARTIFACT	GB 48857448
B	-10	GLN	-	CLONING ARTIFACT	GB 48857448
B	-9	SER	-	CLONING ARTIFACT	GB 48857448
B	-8	THR	-	CLONING ARTIFACT	GB 48857448
B	-7	SER	-	CLONING ARTIFACT	GB 48857448
B	-6	LEU	-	CLONING ARTIFACT	GB 48857448
B	-5	TYR	-	CLONING ARTIFACT	GB 48857448
B	-4	LYS	-	CLONING ARTIFACT	GB 48857448
B	-3	LYS	-	CLONING ARTIFACT	GB 48857448
B	-2	ALA	-	CLONING ARTIFACT	GB 48857448
B	-1	GLY	-	CLONING ARTIFACT	GB 48857448
B	0	LEU	-	CLONING ARTIFACT	GB 48857448
B	1	MSE	MET	MODIFIED RESIDUE	GB 48857448
B	23	MSE	MET	MODIFIED RESIDUE	GB 48857448

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	27	MSE	MET	MODIFIED RESIDUE	GB 48857448
B	75	MSE	MET	MODIFIED RESIDUE	GB 48857448
B	94	MSE	MET	MODIFIED RESIDUE	GB 48857448

- Molecule 2 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total X 1 1	0	0
2	A	3	Total X 3 3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	46	Total O 46 46	0	0
3	B	44	Total O 44 44	0	0

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: Conserved hypothetical protein

Chain A:

- Molecule 1: Conserved hypothetical protein

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	49.91Å 36.79Å 61.82Å 90.00° 111.93° 90.00°	Depositor
Resolution (Å)	30.96 – 1.80 30.97 – 1.71	Depositor EDS
% Data completeness (in resolution range)	100.0 (30.96-1.80) 73.3 (30.97-1.71)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.42 (at 1.71Å)	Xtriage
Refinement program	REFMAC refmac_5.2.0005	Depositor
R, R_{free}	0.225 , 0.256 0.246 , 0.276	Depositor DCC
R_{free} test set	875 reflections (5.23%)	wwPDB-VP
Wilson B-factor (Å^2)	22.3	Xtriage
Anisotropy	0.772	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 39.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.041 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	1439	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.11% 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

5.1 Standard geometry

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

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	0.73	1/670 (0.1%)	0.72	0/897
1	B	0.64	0/669	0.76	0/897
All	All	0.69	1/1339 (0.1%)	0.74	0/1794

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	73	VAL	CB-CG1	-5.18	1.42	1.52

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	673	0	678	4	0
1	B	672	0	668	2	0
2	A	3	0	0	0	0
2	B	1	0	0	0	0
3	A	46	0	0	0	0
3	B	44	0	0	0	0
All	All	1439	0	1346	4	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 1.

All (4) 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:73:VAL:HG11	1:B:63:ILE:O	2.11	0.51
1:A:53:THR:HG23	1:A:59:LYS:HG2	1.95	0.48
1:A:76:LEU:O	1:A:80:ILE:HG12	2.17	0.45
1:A:68:VAL:HG21	1:B:73:VAL:HG12	1.99	0.44

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	89/143 (62%)	88 (99%)	1 (1%)	0	100	100
1	B	91/143 (64%)	90 (99%)	1 (1%)	0	100	100
All	All	180/286 (63%)	178 (99%)	2 (1%)	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	69/107 (64%)	68 (99%)	1 (1%)	67	59

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	66/107 (62%)	65 (98%)	1 (2%)	65 56
All	All	135/214 (63%)	133 (98%)	2 (2%)	65 56

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	89	ARG
1	B	42	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 4 ligands modelled in this entry, 4 are unknown - 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

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.