Full wwPDB X-ray Structure Validation Report ⓘ

Feb 14, 2024 – 09:50 AM EST

PDB ID	:	3KCP
Title	:	Crystal structure of interacting Clostridium thermocellum multimodular components
Authors	:	Adams, J.J.; Currie, M.A.; Bayer, E.A.; Jia, Z.; Smith, S.P.
Deposited on	:	2009-10-21
Resolution	:	1.94 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

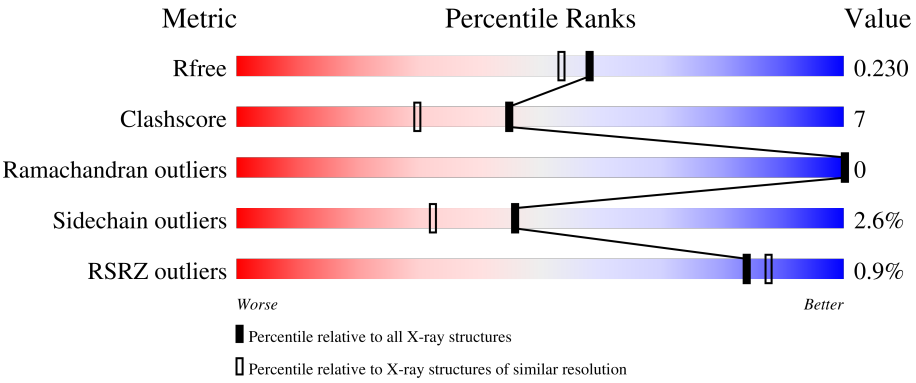
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

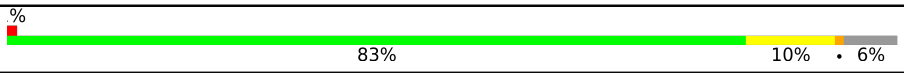
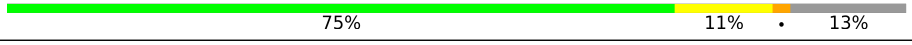
The reported resolution of this entry is 1.94 Å.

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	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	321	
2	B	187	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3789 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 Cellulosomal-scaffolding protein A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	303	Total	C	N	O	S	0	0	0
			2290	1464	363	457	6			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP Q06851
A	314	LEU	-	expression tag	UNP Q06851
A	315	GLU	-	expression tag	UNP Q06851
A	316	HIS	-	expression tag	UNP Q06851
A	317	HIS	-	expression tag	UNP Q06851
A	318	HIS	-	expression tag	UNP Q06851
A	319	HIS	-	expression tag	UNP Q06851
A	320	HIS	-	expression tag	UNP Q06851
A	321	HIS	-	expression tag	UNP Q06851

- Molecule 2 is a protein called Cellulosome anchoring protein, cohesin region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	163	Total	C	N	O	S	0	0	0
			1252	805	201	244	2			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	423	MET	-	expression tag	UNP A3DF10
B	424	ARG	-	expression tag	UNP A3DF10
B	425	GLY	-	expression tag	UNP A3DF10
B	426	SER	-	expression tag	UNP A3DF10
B	427	HIS	-	expression tag	UNP A3DF10
B	428	HIS	-	expression tag	UNP A3DF10
B	429	HIS	-	expression tag	UNP A3DF10

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	430	HIS	-	expression tag	UNP A3DF10
B	431	HIS	-	expression tag	UNP A3DF10
B	432	HIS	-	expression tag	UNP A3DF10
B	433	THR	-	expression tag	UNP A3DF10
B	434	ASP	-	expression tag	UNP A3DF10
B	435	LEU	-	expression tag	UNP A3DF10

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Ca 2 2	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Cl 1 1	0	0

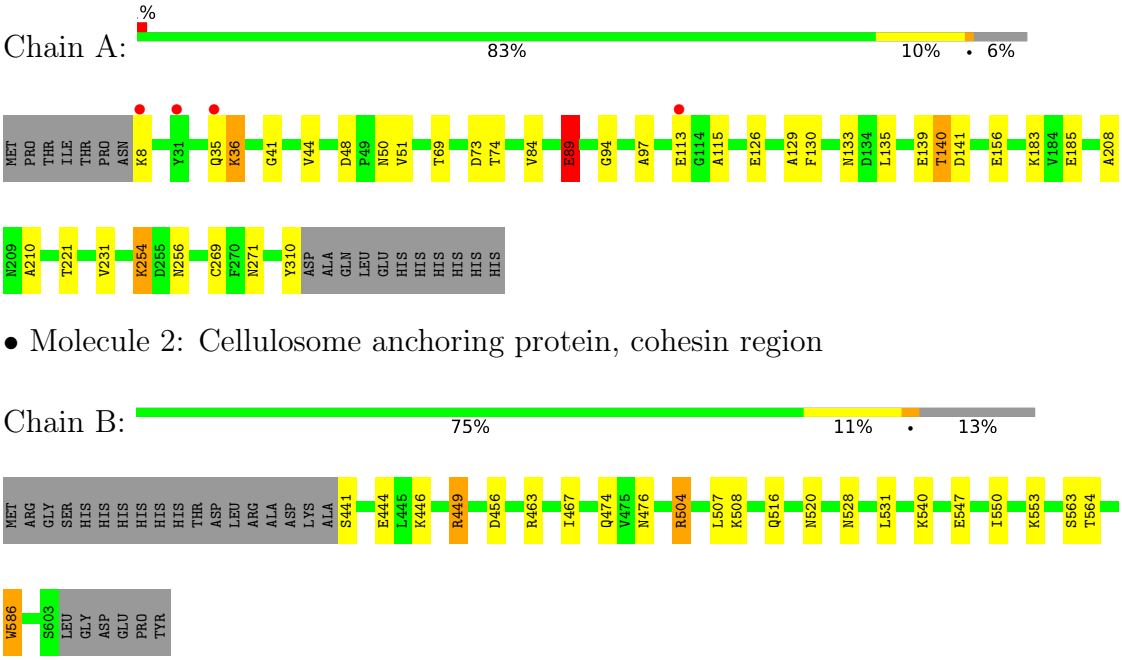
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	150	Total O 150 150	0	0
5	B	94	Total O 94 94	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: Cellulosomal-scaffolding protein A



• Molecule 2: Cellulosome anchoring protein, cohesin region

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	169.98Å 58.02Å 56.15Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.49 – 1.94 40.54 – 1.94	Depositor EDS
% Data completeness (in resolution range)	98.7 (42.49-1.94) 98.7 (40.54-1.94)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	1.48 (at 1.94Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.206 , 0.231 0.206 , 0.230	Depositor DCC
R_{free} test set	2085 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	34.7	Xtriage
Anisotropy	0.095	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 42.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.020 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3789	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.64% 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: CA, CL

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	0.93	3/2333 (0.1%)	0.81	0/3179
2	B	0.85	0/1276	0.86	3/1728 (0.2%)
All	All	0.90	3/3609 (0.1%)	0.83	3/4907 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	89	GLU	CB-CG	8.65	1.68	1.52
1	A	269	CYS	CB-SG	-7.23	1.70	1.82
1	A	208	ALA	CA-CB	5.18	1.63	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	449	ARG	NE-CZ-NH2	-8.94	115.83	120.30
2	B	504	ARG	NE-CZ-NH1	-8.89	115.85	120.30
2	B	504	ARG	NE-CZ-NH2	5.82	123.21	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	2290	0	2265	29	0
2	B	1252	0	1256	19	0
3	A	2	0	0	0	0
4	A	1	0	0	0	0
5	A	150	0	0	13	0
5	B	94	0	0	5	0
All	All	3789	0	3521	46	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.

All (46) 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:310:TYR:C	5:A:462:HOH:O	1.64	1.34
2:B:446:LYS:HE3	5:B:179:HOH:O	1.65	0.96
2:B:476:ASN:HD22	2:B:528:ASN:HD22	1.20	0.87
1:A:156:GLU:HG3	5:A:386:HOH:O	1.78	0.84
2:B:449:ARG:CD	5:B:48:HOH:O	2.25	0.83
2:B:449:ARG:HD3	5:B:48:HOH:O	1.78	0.82
2:B:504:ARG:NH1	2:B:507:LEU:O	2.15	0.79
1:A:139:GLU:HG3	5:A:425:HOH:O	1.86	0.75
1:A:73:ASP:HB3	5:A:431:HOH:O	1.88	0.73
1:A:271:ASN:HD21	2:B:474:GLN:HE22	1.37	0.70
1:A:221:THR:H	1:A:254:LYS:HZ1	1.41	0.66
2:B:516:GLN:HB3	2:B:531:LEU:HD12	1.77	0.66
1:A:156:GLU:CG	5:A:386:HOH:O	2.39	0.65
1:A:35:GLN:HG2	5:A:387:HOH:O	1.98	0.64
2:B:520:ASN:HD22	2:B:528:ASN:H	1.47	0.63
2:B:449:ARG:HD2	5:B:48:HOH:O	1.94	0.63
1:A:139:GLU:CD	1:A:139:GLU:H	2.01	0.62
2:B:441:SER:N	5:B:157:HOH:O	2.31	0.62
1:A:254:LYS:HE2	5:A:416:HOH:O	2.00	0.62
2:B:476:ASN:ND2	2:B:528:ASN:HD22	1.96	0.61
2:B:449:ARG:HH22	2:B:456:ASP:HB3	1.66	0.60
2:B:504:ARG:HH12	2:B:507:LEU:C	2.04	0.56
1:A:41:GLY:HA2	1:A:129:ALA:O	2.06	0.56
1:A:97:ALA:HB2	1:A:133:ASN:HD21	1.72	0.55
1:A:74:THR:HA	1:A:84:VAL:O	2.09	0.53
1:A:156:GLU:HB2	5:A:398:HOH:O	2.07	0.53
1:A:89:GLU:OE2	1:A:94:GLY:N	2.29	0.53
2:B:550:ILE:HD13	2:B:553:LYS:HE3	1.92	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:467:ILE:O	2:B:547:GLU:HG2	2.11	0.51
1:A:8:LYS:HE3	1:A:141:ASP:OD1	2.11	0.51
1:A:271:ASN:HD21	2:B:474:GLN:NE2	2.09	0.49
1:A:48:ASP:OD1	1:A:50:ASN:HB2	2.13	0.49
1:A:130:PHE:HD1	1:A:140:THR:HG21	1.79	0.48
2:B:444:GLU:OE2	2:B:446:LYS:HE2	2.14	0.47
1:A:44:VAL:HB	1:A:126:GLU:HB3	2.00	0.44
1:A:210:ALA:HA	1:A:231:VAL:HG13	2.00	0.43
1:A:256:ASN:ND2	5:A:331:HOH:O	2.51	0.42
2:B:540:LYS:HA	2:B:586:TRP:CE3	2.54	0.42
1:A:183:LYS:HE2	1:A:185:GLU:CG	2.49	0.42
1:A:73:ASP:HB2	5:A:406:HOH:O	2.18	0.42
1:A:36:LYS:HB2	5:A:443:HOH:O	2.19	0.42
1:A:221:THR:H	1:A:254:LYS:NZ	2.15	0.41
2:B:504:ARG:O	2:B:508:LYS:HD2	2.20	0.41
1:A:69:THR:HG22	5:A:326:HOH:O	2.21	0.41
1:A:8:LYS:N	5:A:401:HOH:O	2.55	0.40
1:A:51:VAL:HA	1:A:115:ALA:HB1	2.04	0.40

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	301/321 (94%)	294 (98%)	7 (2%)	0	100	100
2	B	161/187 (86%)	158 (98%)	3 (2%)	0	100	100
All	All	462/508 (91%)	452 (98%)	10 (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	249/266 (94%)	243 (98%)	6 (2%)	49	36
2	B	136/156 (87%)	132 (97%)	4 (3%)	42	28
All	All	385/422 (91%)	375 (97%)	10 (3%)	46	32

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	36	LYS
1	A	89	GLU
1	A	113	GLU
1	A	135	LEU
1	A	140	THR
1	A	254	LYS
2	B	463	ARG
2	B	563	SER
2	B	564	THR
2	B	586	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	133	ASN
1	A	256	ASN
1	A	271	ASN
2	B	476	ASN
2	B	520	ASN
2	B	582	GLN

5.3.3 RNA (i)

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 monosaccharides in this entry.

5.6 Ligand geometry

Of 3 ligands modelled in this entry, 3 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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	303/321 (94%)	0.27	4 (1%) 77 81	22, 33, 47, 60	0
2	B	163/187 (87%)	0.24	0 100 100	23, 34, 46, 50	0
All	All	466/508 (91%)	0.26	4 (0%) 84 87	22, 33, 47, 60	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	35	GLN	2.8
1	A	8	LYS	2.4
1	A	113	GLU	2.2
1	A	31	TYR	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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. The B-factors column lists the minimum, median, 95^{th} 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	B-factors(Å^2)	Q<0.9
4	CL	A	703	1/1	0.93	0.07	56,56,56,56	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CA	A	702	1/1	0.97	0.17	35,35,35,35	0
3	CA	A	701	1/1	0.99	0.14	25,25,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.