Full wwPDB X-ray Structure Validation Report ⓘ

Nov 14, 2023 – 03:17 pm GMT

PDB ID	:	8BN7
Title	:	CjCel5C endo-glucanase
Authors	:	McGregor, N.G.S.; Davies, G.J.
Deposited on	:	2022-11-13
Resolution	:	2.18 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:

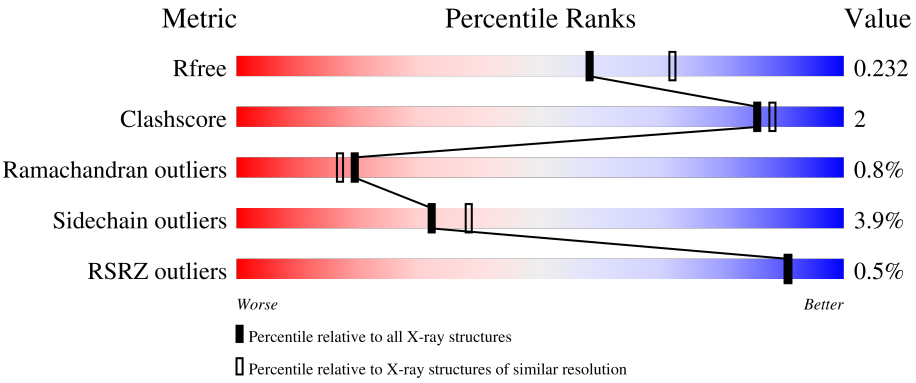
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.18 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	6864 (2.20-2.16)
Clashscore	141614	7689 (2.20-2.16)
Ramachandran outliers	138981	7564 (2.20-2.16)
Sidechain outliers	138945	7564 (2.20-2.16)
RSRZ outliers	127900	6738 (2.20-2.16)

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	AAA	346	89% 6%
1	BBB	346	87% 8%
1	CCC	346	% 87% 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 16101 atoms, of which 7772 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 Cellulase, putative, cel5C.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	CCC	333	Total	C	H	N	O	S	70	3	0
			5330	1751	2604	477	490	8			
1	AAA	331	Total	C	H	N	O	S	64	0	0
			5299	1741	2589	471	490	8			
1	BBB	334	Total	C	H	N	O	S	70	0	0
			5289	1742	2579	470	490	8			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
CCC	1	MET	-	initiating methionine	UNP B3PF55
CCC	339	LEU	-	expression tag	UNP B3PF55
CCC	340	GLU	-	expression tag	UNP B3PF55
CCC	341	HIS	-	expression tag	UNP B3PF55
CCC	342	HIS	-	expression tag	UNP B3PF55
CCC	343	HIS	-	expression tag	UNP B3PF55
CCC	344	HIS	-	expression tag	UNP B3PF55
CCC	345	HIS	-	expression tag	UNP B3PF55
CCC	346	HIS	-	expression tag	UNP B3PF55
AAA	1	MET	-	initiating methionine	UNP B3PF55
AAA	339	LEU	-	expression tag	UNP B3PF55
AAA	340	GLU	-	expression tag	UNP B3PF55
AAA	341	HIS	-	expression tag	UNP B3PF55
AAA	342	HIS	-	expression tag	UNP B3PF55
AAA	343	HIS	-	expression tag	UNP B3PF55
AAA	344	HIS	-	expression tag	UNP B3PF55
AAA	345	HIS	-	expression tag	UNP B3PF55
AAA	346	HIS	-	expression tag	UNP B3PF55
BBB	1	MET	-	initiating methionine	UNP B3PF55
BBB	339	LEU	-	expression tag	UNP B3PF55
BBB	340	GLU	-	expression tag	UNP B3PF55
BBB	341	HIS	-	expression tag	UNP B3PF55
BBB	342	HIS	-	expression tag	UNP B3PF55

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	343	HIS	-	expression tag	UNP B3PF55
BBB	344	HIS	-	expression tag	UNP B3PF55
BBB	345	HIS	-	expression tag	UNP B3PF55
BBB	346	HIS	-	expression tag	UNP B3PF55

- Molecule 2 is water.

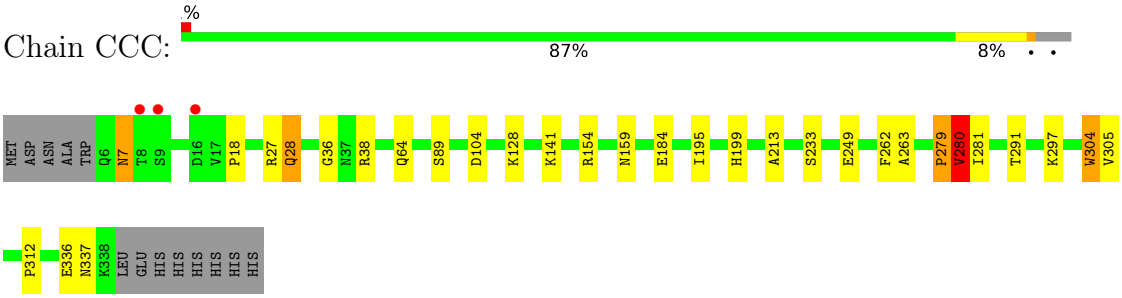
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	CCC	60	Total O 61 61	0	1
2	AAA	51	Total O 53 53	0	2
2	BBB	69	Total O 69 69	0	0

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.

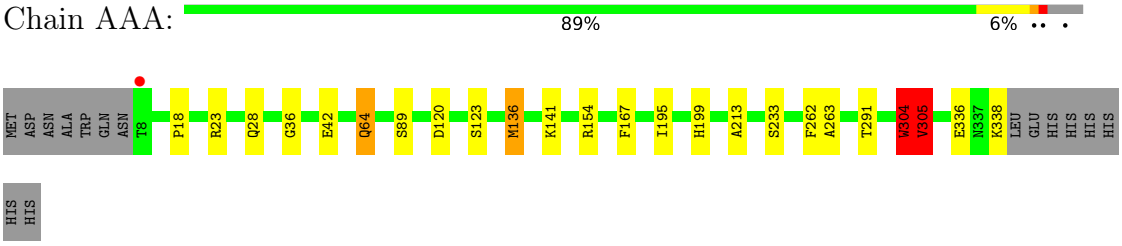
• Molecule 1: Cellulase, putative, cel5C

Chain CCC:



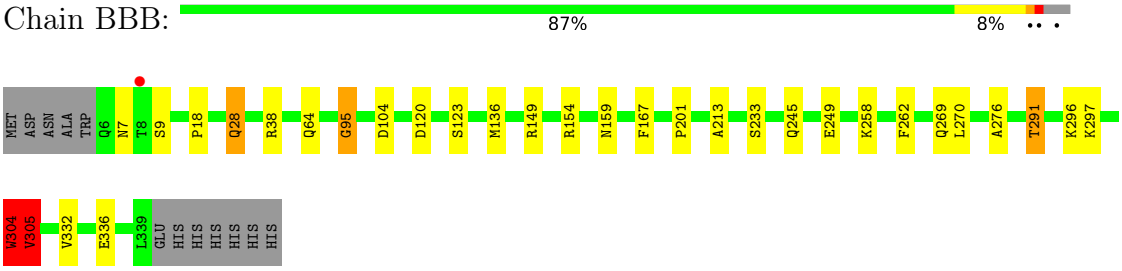
• Molecule 1: Cellulase, putative, cel5C

Chain AAA:



• Molecule 1: Cellulase, putative, cel5C

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	126.33Å 175.71Å 115.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	102.57 – 2.18 102.57 – 2.18	Depositor EDS
% Data completeness (in resolution range)	90.6 (102.57-2.18) 90.6 (102.57-2.18)	Depositor EDS
R_{merge}	0.26	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.76 (at 2.18Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.187 , 0.226 0.195 , 0.232	Depositor DCC
R_{free} test set	2991 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	36.9	Xtriage
Anisotropy	0.068	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 37.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	16101	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.60% 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

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	AAA	0.74	0/2797	0.88	2/3802 (0.1%)
1	BBB	0.73	0/2797	0.88	3/3805 (0.1%)
1	CCC	0.73	0/2822	0.87	4/3838 (0.1%)
All	All	0.74	0/8416	0.88	9/11445 (0.1%)

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	AAA	0	2
1	BBB	0	3
1	CCC	0	2
All	All	0	7

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	BBB	149	ARG	NE-CZ-NH2	-6.55	117.02	120.30
1	CCC	280	VAL	N-CA-CB	-6.02	98.26	111.50
1	AAA	304	TRP	C-N-CA	5.91	136.48	121.70
1	BBB	304	TRP	C-N-CA	5.90	136.46	121.70
1	BBB	305	VAL	N-CA-CB	5.88	124.45	111.50
1	AAA	305	VAL	N-CA-CB	5.87	124.42	111.50
1	CCC	279	PRO	N-CA-CB	-5.62	96.41	102.60
1	CCC	304	TRP	C-N-CA	5.26	134.84	121.70
1	CCC	89	SER	N-CA-CB	-5.14	102.79	110.50

There are no chirality outliers.

All (7) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	304	TRP	Peptide,Mainchain
1	BBB	304	TRP	Peptide,Mainchain
1	BBB	95	GLY	Peptide
1	CCC	304	TRP	Peptide,Mainchain

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	AAA	2710	2589	2579	10	0
1	BBB	2710	2579	2556	15	0
1	CCC	2726	2604	2585	12	0
2	AAA	53	0	0	1	0
2	BBB	69	0	0	0	0
2	CCC	61	0	0	1	0
All	All	8329	7772	7720	35	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 2.

All (35) 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:AAA:304:TRP:CD1	1:AAA:305:VAL:HG22	2.20	0.77
1:BBB:304:TRP:CD1	1:BBB:305:VAL:HG22	2.19	0.77
1:CCC:18:PRO:O	1:CCC:154:ARG:NH2	2.28	0.67
1:BBB:18:PRO:O	1:BBB:154:ARG:NH2	2.29	0.66
1:AAA:18:PRO:O	1:AAA:154:ARG:NH2	2.28	0.65
1:AAA:199:HIS:HE1	2:AAA:443:HOH:O	1.88	0.56
1:BBB:249:GLU:OE1	1:BBB:297:LYS:HE2	2.11	0.50
1:CCC:199:HIS:HE1	2:CCC:444:HOH:O	1.94	0.50
1:CCC:128:LYS:HE3	1:CCC:184:GLU:OE2	2.13	0.48
1:AAA:42:GLU:HG2	1:BBB:296:LYS:HE2	1.96	0.48
1:AAA:195:ILE:O	1:AAA:199:HIS:HD2	1.97	0.48
1:CCC:28:GLN:HE21	1:CCC:28:GLN:HA	1.80	0.45
1:BBB:28:GLN:HE22	1:BBB:159:ASN:H	1.65	0.45
1:CCC:195:ILE:O	1:CCC:199:HIS:HD2	1.99	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:CCC:28:GLN:HE22	1:CCC:159:ASN:H	1.65	0.44
1:AAA:123:SER:HB3	1:AAA:167:PHE:CE2	2.53	0.44
1:AAA:213:ALA:O	1:AAA:233:SER:HA	2.18	0.43
1:BBB:213:ALA:O	1:BBB:233:SER:HA	2.19	0.43
1:CCC:233:SER:O	1:CCC:263:ALA:HA	2.18	0.43
1:AAA:233:SER:O	1:AAA:263:ALA:HA	2.19	0.43
1:CCC:213:ALA:O	1:CCC:233:SER:HA	2.19	0.43
1:BBB:28:GLN:HE21	1:BBB:28:GLN:HA	1.83	0.42
1:CCC:281:ILE:HG22	1:BBB:201:PRO:HB3	2.01	0.42
1:BBB:9:SER:HB2	1:BBB:95:GLY:HA2	2.02	0.42
1:BBB:123:SER:HB3	1:BBB:167:PHE:CE2	2.55	0.42
1:BBB:104:ASP:OD2	1:BBB:154:ARG:NH1	2.53	0.42
1:BBB:245:GLN:HE22	1:BBB:296:LYS:NZ	2.18	0.42
1:BBB:291:THR:HG21	1:BBB:332:VAL:HG21	2.02	0.42
1:AAA:64:GLN:HE21	1:AAA:64:GLN:HB2	1.63	0.41
1:AAA:89:SER:HB2	1:AAA:136:MET:O	2.20	0.41
1:BBB:269:GLN:NE2	1:BBB:276:ALA:HA	2.36	0.41
1:CCC:249:GLU:OE1	1:CCC:297:LYS:HE2	2.21	0.41
1:CCC:280:VAL:HG21	1:CCC:312:PRO:HD3	2.03	0.40
1:CCC:104:ASP:OD2	1:CCC:154:ARG:NH1	2.54	0.40
1:BBB:269:GLN:HE21	1:BBB:276:ALA:HA	1.85	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	AAA	329/346 (95%)	314 (95%)	13 (4%)	2 (1%)	25 24
1	BBB	332/346 (96%)	311 (94%)	19 (6%)	2 (1%)	25 24
1	CCC	334/346 (96%)	313 (94%)	17 (5%)	4 (1%)	13 9
All	All	995/1038 (96%)	938 (94%)	49 (5%)	8 (1%)	19 17

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	CCC	305	VAL
1	AAA	305	VAL
1	BBB	305	VAL
1	CCC	36	GLY
1	AAA	36	GLY
1	CCC	7	ASN
1	BBB	7	ASN
1	CCC	337	ASN

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	AAA	283/297 (95%)	272 (96%)	11 (4%)	32	38
1	BBB	279/297 (94%)	268 (96%)	11 (4%)	32	38
1	CCC	282/297 (95%)	271 (96%)	11 (4%)	32	38
All	All	844/891 (95%)	811 (96%)	33 (4%)	32	38

All (33) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	CCC	7	ASN
1	CCC	27	ARG
1	CCC	28	GLN
1	CCC	38	ARG
1	CCC	64	GLN
1	CCC	141	LYS
1	CCC	262	PHE
1	CCC	279	PRO
1	CCC	280	VAL
1	CCC	291	THR
1	CCC	336	GLU
1	AAA	23	ARG
1	AAA	28	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	AAA	64	GLN
1	AAA	120	ASP
1	AAA	136	MET
1	AAA	141	LYS
1	AAA	262	PHE
1	AAA	291	THR
1	AAA	305	VAL
1	AAA	336	GLU
1	AAA	338	LYS
1	BBB	28	GLN
1	BBB	38	ARG
1	BBB	64	GLN
1	BBB	120	ASP
1	BBB	136	MET
1	BBB	258	LYS
1	BBB	262	PHE
1	BBB	270	LEU
1	BBB	291	THR
1	BBB	305	VAL
1	BBB	336	GLU

Sometimes 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 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

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	AAA	331/346 (95%)	-0.28	1 (0%)	94	94	26, 38, 59, 120	0
1	BBB	334/346 (96%)	-0.19	1 (0%)	94	94	28, 42, 70, 110	0
1	CCC	333/346 (96%)	-0.20	3 (0%)	84	84	26, 38, 63, 119	0
All	All	998/1038 (96%)	-0.22	5 (0%)	91	91	26, 39, 65, 120	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	CCC	8	THR	5.0
1	AAA	8	THR	4.8
1	CCC	16	ASP	2.8
1	CCC	9	SER	2.3
1	BBB	8	THR	2.2

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.