



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:17 PM GMT

PDB ID : 4HU0  
Title : Crystal Structure of a metagenome-derived cellulase Cel5A in complex with cellotetraose  
Authors : Zhuang, N.; Lee, K.H.  
Deposited on : 2012-11-02  
Resolution : 2.38 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](mailto: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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

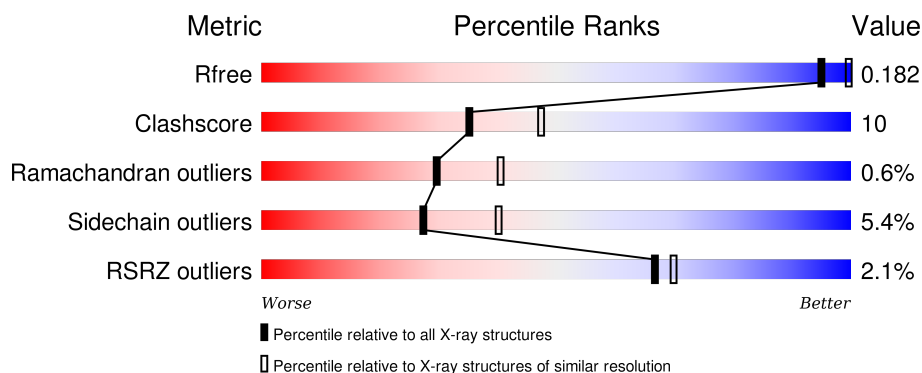
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.38 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))
$R_{free}$	91344	4019 (2.40-2.36)
Clashscore	102246	4595 (2.40-2.36)
Ramachandran outliers	100387	4520 (2.40-2.36)
Sidechain outliers	100360	4522 (2.40-2.36)
RSRZ outliers	91569	4034 (2.40-2.36)

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  $\geq 3$ , 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	359	<div> <div>2%</div> <div>70%</div> <div>18%</div> <div>8%</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
3	GOL	A	402	-	-	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	GOL	A	407	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3070 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.

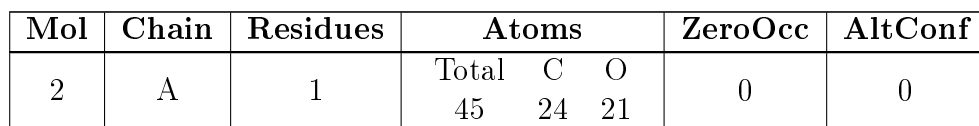
- Molecule 1 is a protein called Cellulase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	330	Total	C	N	O	Se	0	2	0
			2718	1755	470	489	4			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	5	MSE	-	EXPRESSION TAG	UNP I6PLH5
A	6	GLY	-	EXPRESSION TAG	UNP I6PLH5
A	7	SER	-	EXPRESSION TAG	UNP I6PLH5
A	8	SER	-	EXPRESSION TAG	UNP I6PLH5
A	9	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	10	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	11	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	12	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	13	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	14	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	15	SER	-	EXPRESSION TAG	UNP I6PLH5
A	16	SER	-	EXPRESSION TAG	UNP I6PLH5
A	17	GLY	-	EXPRESSION TAG	UNP I6PLH5
A	18	LEU	-	EXPRESSION TAG	UNP I6PLH5
A	19	VAL	-	EXPRESSION TAG	UNP I6PLH5
A	20	PRO	-	EXPRESSION TAG	UNP I6PLH5
A	21	ARG	-	EXPRESSION TAG	UNP I6PLH5
A	22	GLY	-	EXPRESSION TAG	UNP I6PLH5
A	23	SER	-	EXPRESSION TAG	UNP I6PLH5
A	24	HIS	-	EXPRESSION TAG	UNP I6PLH5
A	25	MSE	-	EXPRESSION TAG	UNP I6PLH5

- Molecule 2 is SUGAR (BETA-D-GLUCOPYRANOSYL-(1->4)-BETA-D-GLUCOPYRANOSYL-(1->4)-BETA-D-GLUCOPYRANOSYL-(1->4)-BETA-D-GLUCOPYRANOSE) (three-letter code: CTT) (formula: C<sub>24</sub>H<sub>42</sub>O<sub>21</sub>).



- GOL
- 
- Diagram illustrating the structure of Glycerol (1,2,3-propanetriol). The structure shows a zigzag chain of three carbon atoms (C1, C2, C3) and three hydroxyl groups (HO01, OH02, OH03).
- The carbon atoms are labeled C1, C2, and C3. The hydroxyl groups are labeled HO01, OH02, and OH03.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0
3	A	1	Total 6	C 3	O 3	0	0
3	A	1	Total 6	C 3	O 3	0	0



*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		

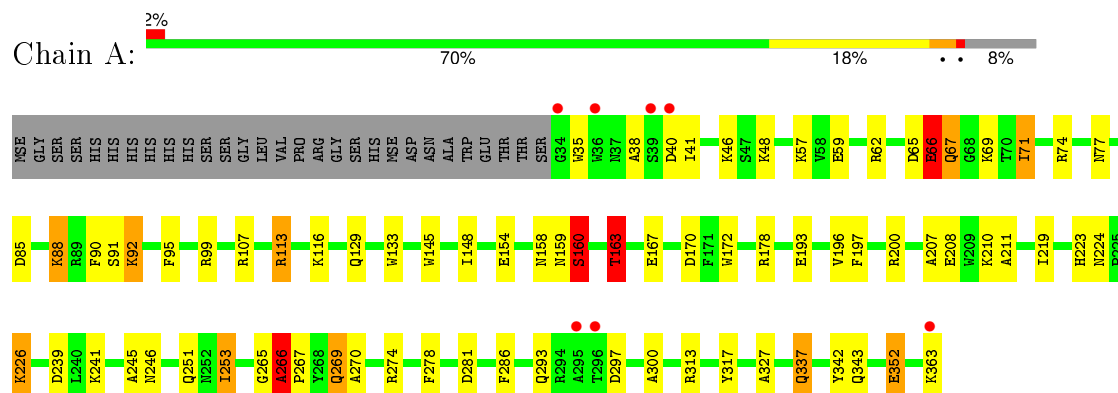
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	271	Total	O	0	0
			271	271		

### 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 errors 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 ( $\text{RSRZ} > 2$ ). 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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	136.92Å 136.92Å 136.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.30 – 2.38 43.30 – 2.38	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.30-2.38) 99.5 (43.30-2.38)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.91 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.152 , 0.188 0.151 , 0.182	Depositor DCC
$R_{free}$ test set	1734 reflections (5.34%)	DCC
Wilson B-factor (Å <sup>2</sup> )	37.9	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 42.4	EDS
Estimated twinning fraction	0.024 for l,-k,h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Outliers	0 of 34226 reflections	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3070	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

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

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	1.53	24/2806 (0.9%)	1.19	25/3809 (0.7%)

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	A	0	1

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	154	GLU	CB-CG	-9.27	1.34	1.52
1	A	266	ALA	N-CA	8.46	1.63	1.46
1	A	352	GLU	CD-OE1	8.19	1.34	1.25
1	A	352	GLU	CG-CD	7.91	1.63	1.51
1	A	154	GLU	CD-OE1	7.27	1.33	1.25

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	62	ARG	NE-CZ-NH2	-12.70	113.95	120.30
1	A	62	ARG	NE-CZ-NH1	11.04	125.82	120.30
1	A	281	ASP	CB-CG-OD1	8.45	125.90	118.30
1	A	92	LYS	CD-CE-NZ	-7.68	94.05	111.70
1	A	297	ASP	CB-CG-OD2	7.00	124.60	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	265	GLY	Peptide

## 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	A	2718	0	2610	51	0
2	A	45	0	42	4	0
3	A	36	0	46	2	0
4	A	271	0	0	14	0
All	All	3070	0	2698	54	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 10.

The worst 5 of 54 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:66[A]:GLU:HG3	1:A:67:GLN:HE22	1.15	1.08
1:A:67:GLN:N	1:A:67:GLN:HE21	1.60	1.00
1:A:269:GLN:H	1:A:269:GLN:HE21	1.10	0.97
1:A:129:GLN:HG2	4:A:679:HOH:O	1.65	0.96
1:A:67:GLN:HE21	1:A:67:GLN:H	0.95	0.94

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	330/359 (92%)	319 (97%)	9 (3%)	2 (1%)	30	40

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	266	ALA
1	A	40	ASP

### 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	282/301 (94%)	266 (94%)	16 (6%)	25	37

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	113	ARG
1	A	160	SER
1	A	253	ILE
1	A	88	LYS
1	A	269	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	224	ASN
1	A	337	GLN
1	A	246	ASN
1	A	223	HIS
1	A	269	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 ⓘ

7 ligands are modelled in this entry.

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$
2	CTT	A	401	-	48,48,48	1.37	7 (14%)	71,71,71	2.61	29 (40%)
3	GOL	A	402	-	5,5,5	1.04	0	5,5,5	0.87	0
3	GOL	A	403	-	5,5,5	1.07	0	5,5,5	1.12	1 (20%)
3	GOL	A	404	-	5,5,5	1.47	1 (20%)	5,5,5	1.74	3 (60%)
3	GOL	A	405	-	5,5,5	0.76	0	5,5,5	1.97	1 (20%)
3	GOL	A	406	-	5,5,5	0.82	0	5,5,5	1.51	1 (20%)
3	GOL	A	407	-	5,5,5	1.13	1 (20%)	5,5,5	1.31	1 (20%)

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
2	CTT	A	401	-	-	0/20/100/100	0/4/4/4
3	GOL	A	402	-	-	0/4/4/4	0/0/0/0
3	GOL	A	403	-	-	0/4/4/4	0/0/0/0
3	GOL	A	404	-	-	0/4/4/4	0/0/0/0
3	GOL	A	405	-	-	0/4/4/4	0/0/0/0
3	GOL	A	406	-	-	0/4/4/4	0/0/0/0
3	GOL	A	407	-	-	0/4/4/4	0/0/0/0

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	404	GOL	O2-C2	-2.74	1.35	1.43
3	A	407	GOL	O2-C2	-2.33	1.36	1.43
2	A	401	CTT	O4D-C1E	2.09	1.47	1.41
2	A	401	CTT	O5B-C1B	2.14	1.47	1.43
2	A	401	CTT	O5D-C1D	2.42	1.48	1.41

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	CTT	O4E-C4E-C5E	-7.47	89.45	109.24
2	A	401	CTT	O6D-C6D-C5D	-4.30	97.13	111.33
3	A	405	GOL	O3-C3-C2	-3.98	90.88	110.18
2	A	401	CTT	O1B-C1B-C2B	-3.97	98.57	109.21
2	A	401	CTT	O5D-C1D-C2D	-3.88	102.32	110.28

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401	CTT	4	0
3	A	402	GOL	1	0
3	A	404	GOL	1	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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	326/359 (90%)	-0.25	7 (2%) 67 70	27, 38, 60, 98	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	363	LYS	4.5
1	A	34	GLY	3.7
1	A	296	THR	2.5
1	A	40	ASP	2.5
1	A	295	ALA	2.2

### 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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
3	GOL	A	402	6/6	0.93	0.14	3.43	38,44,57,60	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	GOL	A	407	6/6	0.94	0.22	3.17	46,60,64,66	0
2	CTT	A	401	45/45	0.96	0.14	1.43	24,44,62,67	8
3	GOL	A	403	6/6	0.74	0.21	0.99	67,74,76,78	0
3	GOL	A	404	6/6	0.95	0.15	0.80	52,55,59,59	0
3	GOL	A	405	6/6	0.88	0.26	-	75,79,85,88	0
3	GOL	A	406	6/6	0.75	0.27	-	78,84,84,85	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.