WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Apr 4, 2022 – 12:33 PM EDT

PDB ID	:	7TOH
Title	:	Crystal structure of carbohydrate esterase PbeAcXE, in complex with MeGlcpA-Xylp
Authors	:	Stogios, P.J.; Skarina, T.; Di Leo, R.; Jurak, E.; Master, E.
Deposited on	:	2022-01-24
Resolution	:	1.26 Å(reported)

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

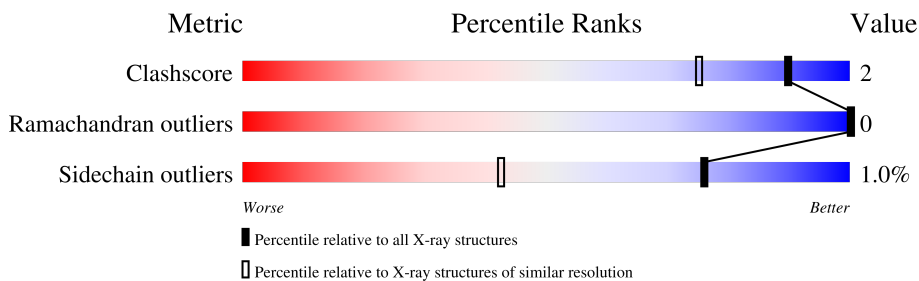|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

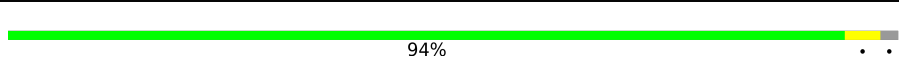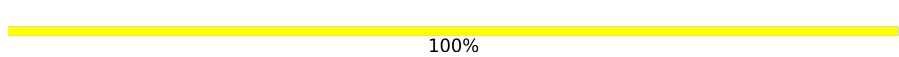The reported resolution of this entry is 1.26 Å.

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	Percentile Ranks	Value
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		1.0%

Worse　　　　　*Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	1060 (1.28-1.24)
Ramachandran outliers	138981	1029 (1.28-1.24)
Sidechain outliers	138945	1028 (1.28-1.24)

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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	386	94%　　　· ·
2	C	2	100%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3713 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 SGNH hydrolase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	379	Total	C	N	O	S	Se	0	10	0
			2964	1875	524	554	2	9			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	initiating methionine	UNP A0A5M4AV20
A	386	GLU	-	expression tag	UNP A0A5M4AV20

- Molecule 2 is an oligosaccharide called 4-O-methyl-alpha-D-glucopyranuronic acid-(1-2)-beta-D-xylopyranose.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	C	2	Total	C	O	0	0	0
			23	12	11			

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	695	Total	O	0	31
			726	726		

3 Residue-property plots ⓘ

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. 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.

Note EDS failed to run properly.

• Molecule 1: SGNH hydrolase

Chain A:
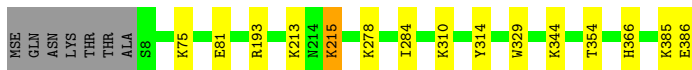
• Molecule 2: 4-O-methyl-alpha-D-glucopyranuronic acid-(1-2)-beta-D-xylopyranose

Chain C:

4 Data and refinement statistics ⓘ

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	57.61Å　42.81Å　76.70Å 90.00°　108.44°　90.00°	Depositor
Resolution (Å)	28.53　–　1.26	Depositor
% Data completeness (in resolution range)	97.4 (28.53-1.26)	Depositor
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.53 (at 1.26Å)	Xtriage
Refinement program	PHENIX 1.20_4459	Depositor
R, R_{free}	0.135　,　0.169	Depositor
Wilson B-factor (Å^2)	16.3	Xtriage
Anisotropy	0.504	Xtriage
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3713	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.80% 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 (i)

5.1 Standard geometry (i)

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

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	0.59	0/3049	0.81	0/4113

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	2964	0	3002	11	0
2	C	23	0	9	0	0
3	A	726	0	0	3	0
All	All	3713	0	3011	11	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 (11) 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:278:LYS:NZ	3:A:401:HOH:O	2.25	0.69

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:310:LYS:HB3	1:A:354[B]:THR:HG22	1.76	0.68
1:A:344:LYS:HE2	3:A:847:HOH:O	2.01	0.60
1:A:385:LYS:HE2	3:A:457[B]:HOH:O	2.15	0.46
1:A:215:LYS:HD2	1:A:215:LYS:O	2.16	0.45
1:A:215:LYS:HE2	1:A:215:LYS:HB3	1.35	0.45
1:A:215:LYS:CE	1:A:215:LYS:C	2.87	0.43
1:A:75:LYS:CG	1:A:81:GLU:HG3	2.49	0.42
1:A:193[B]:ARG:HD3	1:A:366:HIS:CE1	2.56	0.41
1:A:213:LYS:HE2	1:A:213:LYS:HB2	1.63	0.41
1:A:284:ILE:HG23	1:A:329:TRP:CE2	2.55	0.41

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	387/386 (100%)	378 (98%)	9 (2%)	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	324/310 (104%)	321 (99%)	3 (1%)	78 47

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	215	LYS
1	A	314	TYR
1	A	386	GLU

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

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 ⓘ

2 monosaccharides 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	XYP	C	1	2	10,10,10	2.34	2 (20%)	14,14,14	2.43	6 (42%)
2	GCV	C	2	2	10,13,14	1.93	3 (30%)	11,18,20	1.54	3 (27%)

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	XYP	C	1	2	-	-	0/1/1/1
2	GCV	C	2	2	-	0/2/23/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	XYP	O5-C1	5.78	1.51	1.43
2	C	2	GCV	O5-C5	3.91	1.47	1.43
2	C	2	GCV	O5-C1	3.52	1.49	1.43
2	C	1	XYP	O5-C5	3.20	1.48	1.43
2	C	2	GCV	C2-C3	-2.03	1.49	1.52

All (9) bond angle outliers are listed below:

| Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	1	XYP	O3-C3-C4	-5.43	99.59	109.99
2	C	1	XYP	O5-C1-C2	4.12	115.55	109.43
2	C	2	GCV	O5-C1-C2	-2.99	106.16	110.77
2	C	1	XYP	C4-C3-C2	2.82	115.78	110.89
2	C	1	XYP	O5-C5-C4	-2.62	106.73	110.77
2	C	2	GCV	C1-O5-C5	-2.48	107.79	112.17
2	C	1	XYP	O2-C2-C3	2.46	116.03	110.35
2	C	2	GCV	O2-C2-C1	2.16	113.57	109.15
2	C	1	XYP	C5-O5-C1	2.02	116.10	112.71

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

Oligosaccharide Chain C
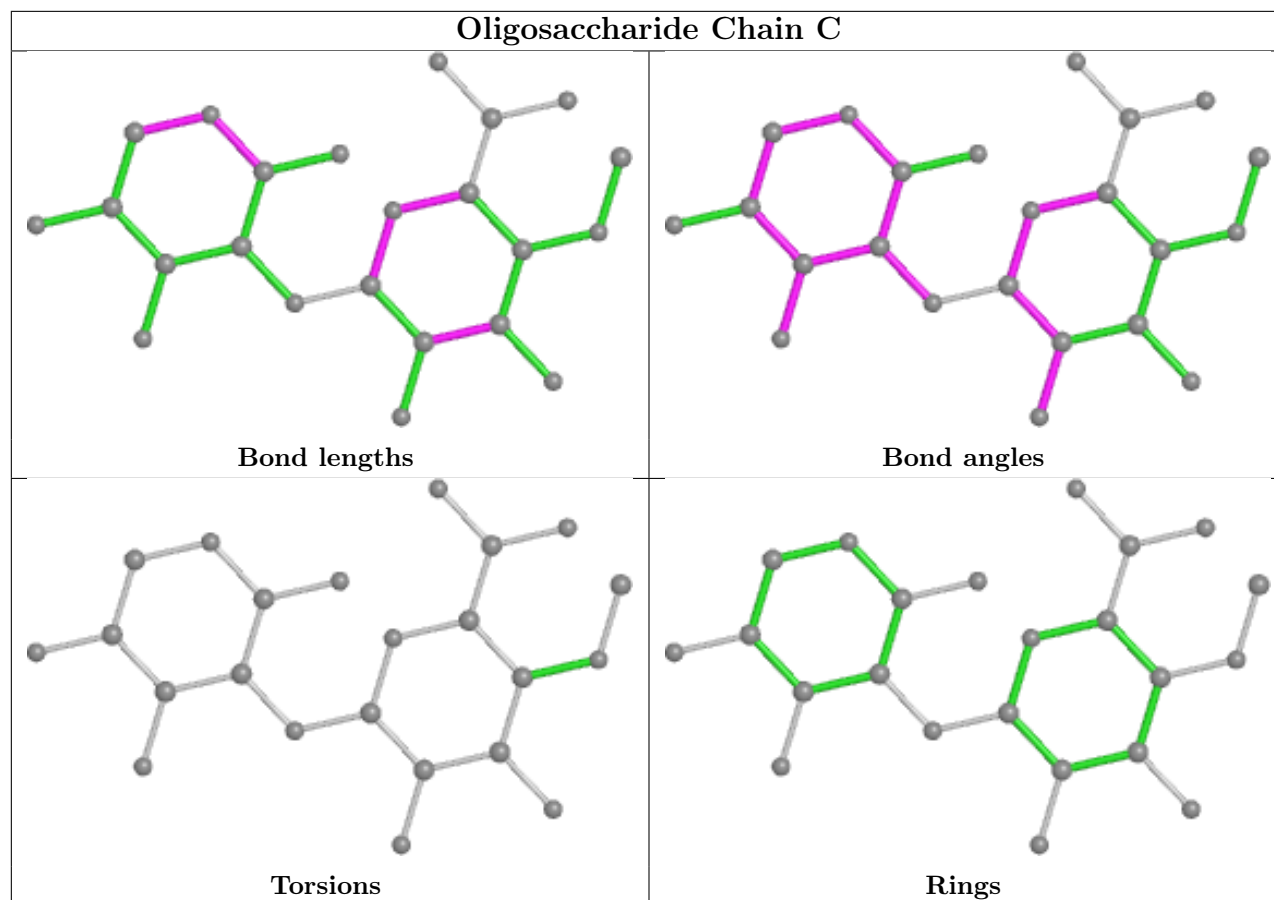
5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS failed to run properly - this section is therefore empty.

6.4 Ligands ⓘ

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers ⓘ

EDS failed to run properly - this section is therefore empty.