



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 31, 2021 – 11:07 pm BST

PDB ID : 7NCX  
Title : Crystal structure of GH30 (double mutant EE) from *Thermothelomyces thermophila*.  
Authors : Dimarogona, M.; Nikolaivits, E.; Topakas, E.; Weiss, M.; Feiler, C.G.  
Deposited on : 2021-01-29  
Resolution : 1.36 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](mailto: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 following versions of software and data (see [references ⓘ](#)) 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	:	2.23.1
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)	:	2.23.1

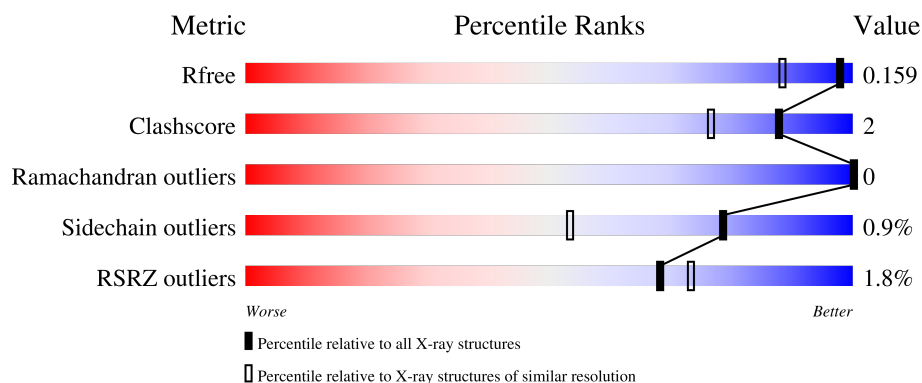
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.36 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	1509 (1.38-1.34)
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	AAA	482	<div> <div>2%</div> <div> <div></div> <div>89%</div> <div>• 8%</div> </div> </div>
2	BBB	6	<div> <div>67%</div> <div>33%</div> </div>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 7489 atoms, of which 3407 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 GH30 family xylanase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	443	Total	C	H	N	O	S	217	17	0
			6756	2156	3296	608	681	15			

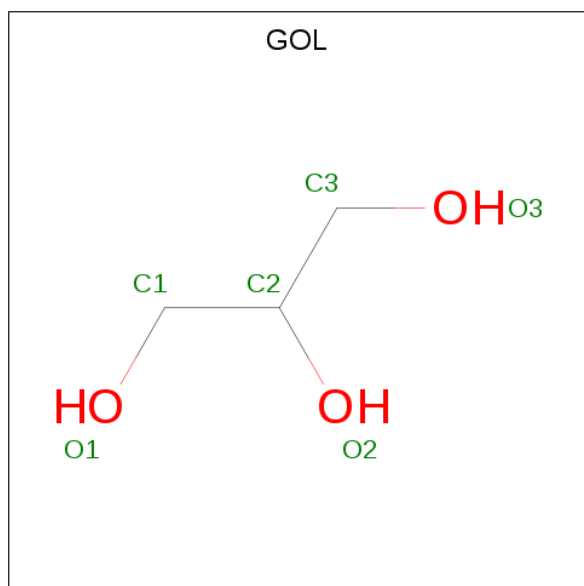
There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	18	SER	-	expression tag	UNP G2Q1N4
AAA	19	MET	-	expression tag	UNP G2Q1N4
AAA	205	ALA	GLU	engineered mutation	UNP G2Q1N4
AAA	295	ALA	GLU	engineered mutation	UNP G2Q1N4
AAA	478	LEU	-	expression tag	UNP G2Q1N4
AAA	479	GLU	-	expression tag	UNP G2Q1N4
AAA	480	GLN	-	expression tag	UNP G2Q1N4
AAA	481	LYS	-	expression tag	UNP G2Q1N4
AAA	482	LEU	-	expression tag	UNP G2Q1N4
AAA	483	ILE	-	expression tag	UNP G2Q1N4
AAA	484	SER	-	expression tag	UNP G2Q1N4
AAA	485	GLU	-	expression tag	UNP G2Q1N4
AAA	486	GLU	-	expression tag	UNP G2Q1N4
AAA	487	ASP	-	expression tag	UNP G2Q1N4
AAA	488	LEU	-	expression tag	UNP G2Q1N4
AAA	489	ASN	-	expression tag	UNP G2Q1N4
AAA	490	SER	-	expression tag	UNP G2Q1N4
AAA	491	ALA	-	expression tag	UNP G2Q1N4
AAA	492	VAL	-	expression tag	UNP G2Q1N4
AAA	493	ASP	-	expression tag	UNP G2Q1N4
AAA	494	HIS	-	expression tag	UNP G2Q1N4
AAA	495	HIS	-	expression tag	UNP G2Q1N4
AAA	496	HIS	-	expression tag	UNP G2Q1N4
AAA	497	HIS	-	expression tag	UNP G2Q1N4
AAA	498	HIS	-	expression tag	UNP G2Q1N4
AAA	499	HIS	-	expression tag	UNP G2Q1N4

- Molecule 2 is an oligosaccharide called alpha-D-mannopyranose-(1-3)-alpha-D-mannopyranose-(1-6)-[alpha-D-mannopyranose-(1-3)]beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	6	Total	C	H	N	O	17	0	0
			139	40	67	2	30			

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	AAA	1	Total	C	H	O	2	0
			14	3	8	3		

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	AAA	1	Total	C	H	O	1	0
			17	4	10	3		

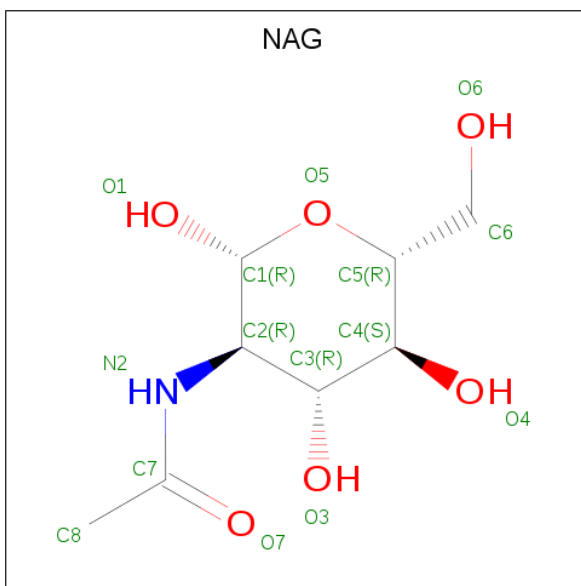
- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
5	AAA	1	Total	C	H	O	1	0
			10	2	6	2		

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:

C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



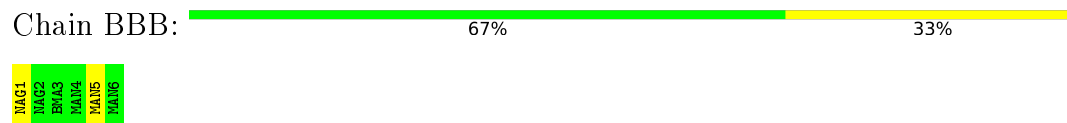
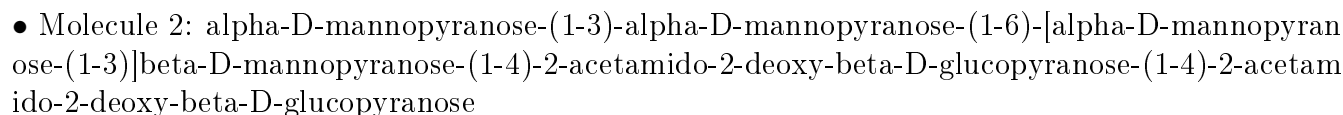
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	AAA	1	Total	C	H	N	O	3	0
			28	8	14	1	5		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	AAA	515	Total	O	0	0
			515	515		



- Molecule 1: GH30 family xylanase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	38.89Å 87.31Å 107.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.70 – 1.36 43.66 – 1.36	Depositor EDS
% Data completeness (in resolution range)	99.6 (43.70-1.36) 99.6 (43.66-1.36)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.75 (at 1.36Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, $R_{free}$	0.122 , 0.158 0.123 , 0.159	Depositor DCC
$R_{free}$ test set	4024 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.7	Xtriage
Anisotropy	0.291	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 44.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	7489	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.45% 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, 0.333 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: PEG, GOL, MAN, BMA, NAG, EDO

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	0.67	0/3543	0.78	0/4833

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

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	3460	3296	3270	14	0
2	BBB	72	67	61	0	0
3	AAA	6	8	8	0	0
4	AAA	7	10	10	1	0
5	AAA	8	12	12	1	0
6	AAA	14	14	13	0	0
7	AAA	515	0	0	2	0
All	All	4082	3407	3374	14	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 (14) 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:342[B]:GLN:NE2	1:AAA:348:SER:OG	2.32	0.62
1:AAA:342[B]:GLN:HG2	1:AAA:349:HIS:CG	2.37	0.59
1:AAA:342[B]:GLN:NE2	1:AAA:349:HIS:ND1	2.49	0.57
1:AAA:342[A]:GLN:HG2	1:AAA:349:HIS:CG	2.45	0.51
1:AAA:169:GLN:NE2	7:AAA:707:HOH:O	2.43	0.50
1:AAA:342[B]:GLN:HE21	1:AAA:349:HIS:CE1	2.29	0.50
1:AAA:449:GLN:NE2	7:AAA:708:HOH:O	2.47	0.48
1:AAA:230:TYR:O	1:AAA:234:GLN:HG2	2.16	0.46
1:AAA:310[A]:TYR:O	1:AAA:310[A]:TYR:CD2	2.71	0.43
1:AAA:374:ARG:HH12	4:AAA:602:PEG:C4	2.32	0.42
1:AAA:51:ARG:HG3	1:AAA:342[B]:GLN:HB3	2.01	0.42
1:AAA:385:VAL:HG12	5:AAA:603:EDO:H11	2.03	0.40
1:AAA:342[B]:GLN:HG2	1:AAA:349:HIS:CB	2.51	0.40
1:AAA:51:ARG:HG3	1:AAA:342[A]:GLN:HB2	2.04	0.40

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	AAA	456/482 (95%)	439 (96%)	17 (4%)	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	AAA	361/379 (95%)	356 (99%)	5 (1%)	67 36

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	194[A]	ASN
1	AAA	194[B]	ASN
1	AAA	227[A]	ARG
1	AAA	227[B]	ARG
1	AAA	252	TRP

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 ⓘ

6 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	NAG	BBB	1	1,2	14,14,15	0.98	1 (7%)	17,19,21	0.77	0
2	NAG	BBB	2	2	14,14,15	0.70	0	17,19,21	0.79	0
2	BMA	BBB	3	2	11,11,12	0.74	0	15,15,17	1.11	0
2	MAN	BBB	4	2	11,11,12	0.92	0	15,15,17	0.82	0
2	MAN	BBB	5	2	11,11,12	0.57	0	15,15,17	0.93	1 (6%)
2	MAN	BBB	6	2	11,11,12	0.65	0	15,15,17	0.99	0

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	NAG	BBB	1	1,2	-	0/6/23/26	0/1/1/1
2	NAG	BBB	2	2	-	0/6/23/26	0/1/1/1
2	BMA	BBB	3	2	-	0/2/19/22	0/1/1/1
2	MAN	BBB	4	2	-	0/2/19/22	0/1/1/1
2	MAN	BBB	5	2	-	0/2/19/22	0/1/1/1
2	MAN	BBB	6	2	-	0/2/19/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	BBB	1	NAG	O5-C1	-3.01	1.38	1.43

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	BBB	5	MAN	C1-O5-C5	2.22	115.19	112.19

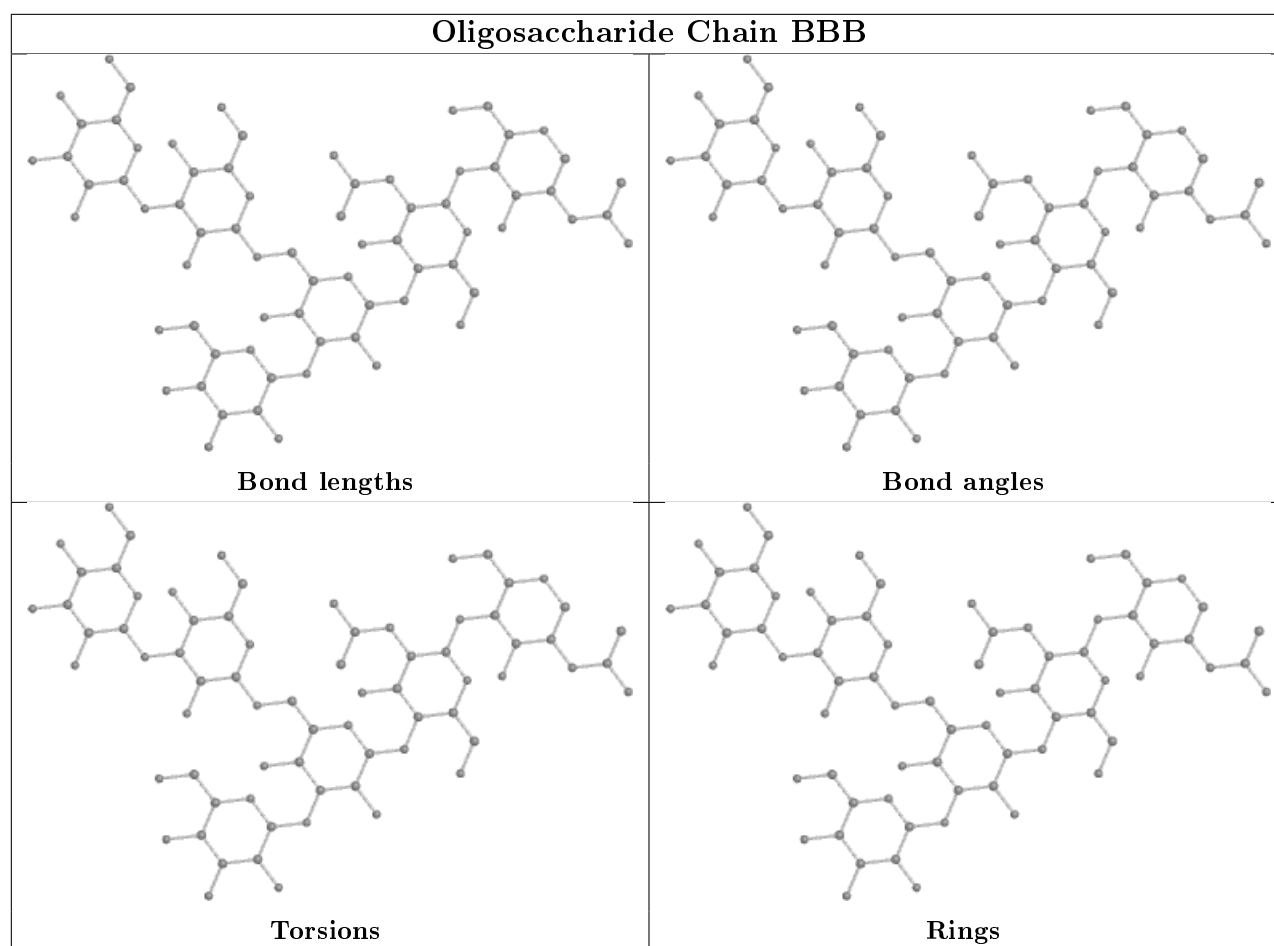
There are no chirality outliers.

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.



## 5.6 Ligand geometry [i](#)

5 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	$\# Z  > 2$
5	EDO	AAA	604	-	3,3,3	0.04	0	2,2,2	0.19	0
4	PEG	AAA	602	-	6,6,6	0.13	0	5,5,5	0.14	0
3	GOL	AAA	601	-	5,5,5	0.15	0	5,5,5	0.41	0
5	EDO	AAA	603	-	3,3,3	0.54	0	2,2,2	0.58	0
6	NAG	AAA	605	1	14,14,15	0.22	0	17,19,21	0.68	0

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
5	EDO	AAA	604	-	-	1/1/1/1	-
4	PEG	AAA	602	-	-	2/4/4/4	-
3	GOL	AAA	601	-	-	4/4/4/4	-
5	EDO	AAA	603	-	-	1/1/1/1	-
6	NAG	AAA	605	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	AAA	601	GOL	O1-C1-C2-C3
5	AAA	604	EDO	O1-C1-C2-O2
3	AAA	601	GOL	O1-C1-C2-O2
5	AAA	603	EDO	O1-C1-C2-O2
3	AAA	601	GOL	O2-C2-C3-O3
4	AAA	602	PEG	C1-C2-O2-C3
3	AAA	601	GOL	C1-C2-C3-O3
4	AAA	602	PEG	O2-C3-C4-O4

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	AAA	602	PEG	1	0
5	AAA	603	EDO	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	AAA	443/482 (91%)	-0.68	8 (1%) 68 73	8, 14, 27, 55	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	24	GLN	3.0
1	AAA	111	PRO	3.0
1	AAA	389	SER	2.3
1	AAA	388	ALA	2.2
1	AAA	300	GLN	2.1
1	AAA	113	ILE	2.1
1	AAA	299	LEU	2.0
1	AAA	474	ALA	2.0

### 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](#)

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<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	B-factors(Å <sup>2</sup> )	Q<0.9
2	MAN	BBB	6	11/12	0.91	0.14	24,29,35,41	4
2	MAN	BBB	5	11/12	0.92	0.13	20,22,28,33	4
2	MAN	BBB	4	11/12	0.92	0.15	20,22,25,27	3
2	NAG	BBB	2	14/15	0.95	0.07	13,17,26,29	2

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	BMA	BBB	3	11/12	0.96	0.10	18,21,27,28	2
2	NAG	BBB	1	14/15	0.98	0.05	10,12,19,19	2

## 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	PEG	AAA	602	7/7	0.65	0.18	33,43,50,51	1
5	EDO	AAA	603	4/4	0.87	0.20	25,33,33,35	1
6	NAG	AAA	605	14/15	0.88	0.19	27,37,56,57	3
3	GOL	AAA	601	6/6	0.93	0.09	25,31,41,42	2
5	EDO	AAA	604	4/4	0.94	0.27	33,37,41,42	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.