



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 20, 2024 – 08:44 pm GMT

PDB ID : 7NWP  
Title : A carbohydrate binding module family 9 (CBM9) from *Caldicellulosiruptor kristjansonii* in complex with cellobiose  
Authors : Krska, D.; Mazurkewich, S.; Navarro Poulsen, J.; Larsbrink, J.; Lo Leggio, L.  
Deposited on : 2021-03-17  
Resolution : 2.16 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

<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  
Mogul : 1.8.4, CSD as541be (2020)  
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

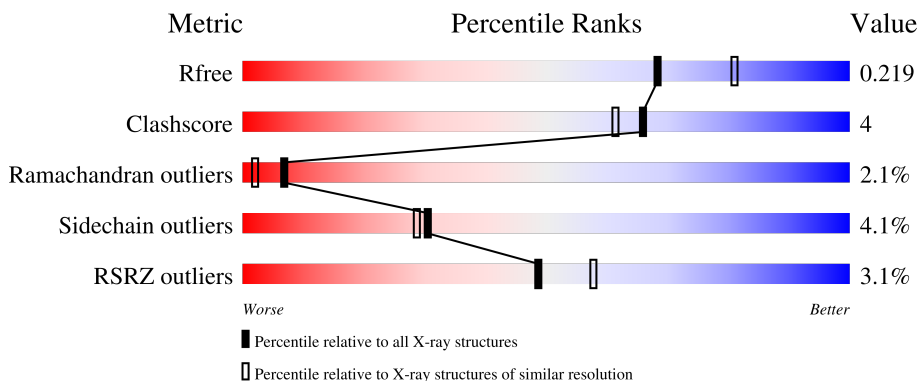
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.16 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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  $\geq 3$ , 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	215	 3% 80% 9% 9%
2	BBB	2	 100%

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 1754 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 Beta-xylanase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	195	1592	1007	271	311	3	0	2	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	1	MET	-	initiating methionine	UNP E4S6E9
AAA	2	GLY	-	expression tag	UNP E4S6E9
AAA	3	SER	-	expression tag	UNP E4S6E9
AAA	4	SER	-	expression tag	UNP E4S6E9
AAA	5	HIS	-	expression tag	UNP E4S6E9
AAA	6	HIS	-	expression tag	UNP E4S6E9
AAA	7	HIS	-	expression tag	UNP E4S6E9
AAA	8	HIS	-	expression tag	UNP E4S6E9
AAA	9	HIS	-	expression tag	UNP E4S6E9
AAA	10	HIS	-	expression tag	UNP E4S6E9
AAA	11	SER	-	expression tag	UNP E4S6E9
AAA	12	SER	-	expression tag	UNP E4S6E9
AAA	13	GLU	-	expression tag	UNP E4S6E9
AAA	14	ASN	-	expression tag	UNP E4S6E9
AAA	15	LEU	-	expression tag	UNP E4S6E9
AAA	16	TYR	-	expression tag	UNP E4S6E9
AAA	17	PHE	-	expression tag	UNP E4S6E9
AAA	18	GLN	-	expression tag	UNP E4S6E9
AAA	19	GLY	-	expression tag	UNP E4S6E9
AAA	20	HIS	-	expression tag	UNP E4S6E9

- Molecule 2 is an oligosaccharide called beta-D-glucopyranose-(1-4)-beta-D-glucopyranose.

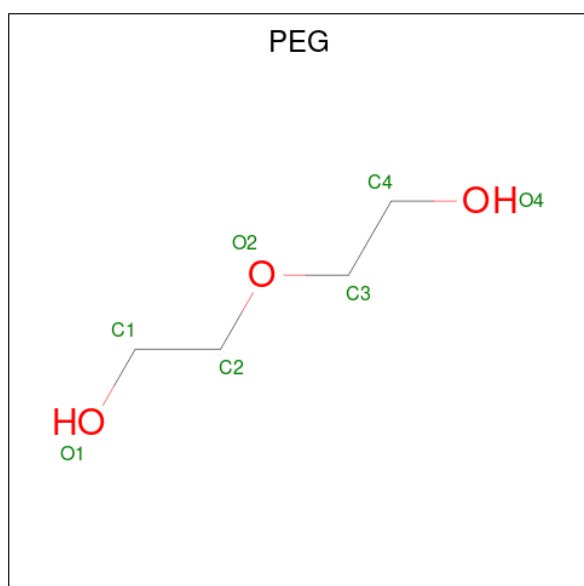


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	BBB	2	Total	C	O	0	0	0
			23	12	11			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	3	Total	Ca	0	0
			3	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	O	0	0
			7	4	3		
4	AAA	1	Total	C	O	0	0
			7	4	3		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAA	1	Total C O 4 2 2	0	0
5	AAA	1	Total C O 4 2 2	0	0
5	AAA	1	Total C O 4 2 2	0	0


- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AAA	110	Total O 110 110	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: Beta-xylanase

Chain AAA: 



P215

- Molecule 2: beta-D-glucopyranose-(1-4)-beta-D-glucopyranose

Chain BBB: 

B6C1  
B6C2

## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	173.32Å 173.32Å 173.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.37 – 2.16 46.32 – 2.16	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.37-2.16) 100.0 (46.32-2.16)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.34 (at 2.16Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.179 , 0.219 0.178 , 0.219	Depositor DCC
$R_{free}$ test set	2400 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.2	Xtrriage
Anisotropy	0.000	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 41.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	1754	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: BGC, EDO, PEG, CA

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	0.49	0/1629	0.86	0/2216

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	AAA	1592	0	1528	13	0
2	BBB	23	0	21	0	0
3	AAA	3	0	0	0	0
4	AAA	14	0	20	0	0
5	AAA	12	0	18	0	0
6	AAA	110	0	0	0	0
All	All	1754	0	1587	13	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 4.

The worst 5 of 13 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:211:VAL:HB	1:AAA:212:PRO:HD3	1.49	0.92
1:AAA:21:LYS:HG2	1:AAA:22:ARG:H	1.38	0.87
1:AAA:211:VAL:CB	1:AAA:212:PRO:HD3	2.22	0.70
1:AAA:211:VAL:HB	1:AAA:212:PRO:CD	2.28	0.60
1:AAA:212:PRO:O	1:AAA:213:ILE:C	2.48	0.52

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	195/215 (91%)	182 (93%)	9 (5%)	4 (2%)	<b>7</b> <b>2</b>

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	211	VAL
1	AAA	213	ILE
1	AAA	212	PRO
1	AAA	209	ALA

### 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	174/190 (92%)	167 (96%)	7 (4%)	<b>31</b> <b>29</b>

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	176	ARG
1	AAA	178	VAL
1	AAA	211	VAL
1	AAA	208	ASN
1	AAA	170	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](#)

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	BGC	BBB	1	2	12,12,12	0.53	0	17,17,17	1.64	4 (23%)
2	BGC	BBB	2	2	11,11,12	0.66	0	15,15,17	2.16	5 (33%)

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	BGC	BBB	1	2	-	2/2/22/22	0/1/1/1
2	BGC	BBB	2	2	-	2/2/19/22	0/1/1/1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	BBB	2	BGC	O2-C2-C3	4.28	118.72	110.14
2	BBB	1	BGC	C1-O5-C5	4.03	121.27	113.66
2	BBB	2	BGC	C1-O5-C5	3.54	116.99	112.19
2	BBB	2	BGC	C1-C2-C3	-3.40	105.49	109.67
2	BBB	1	BGC	O1-C1-O5	-3.02	101.33	110.38

There are no chirality outliers.

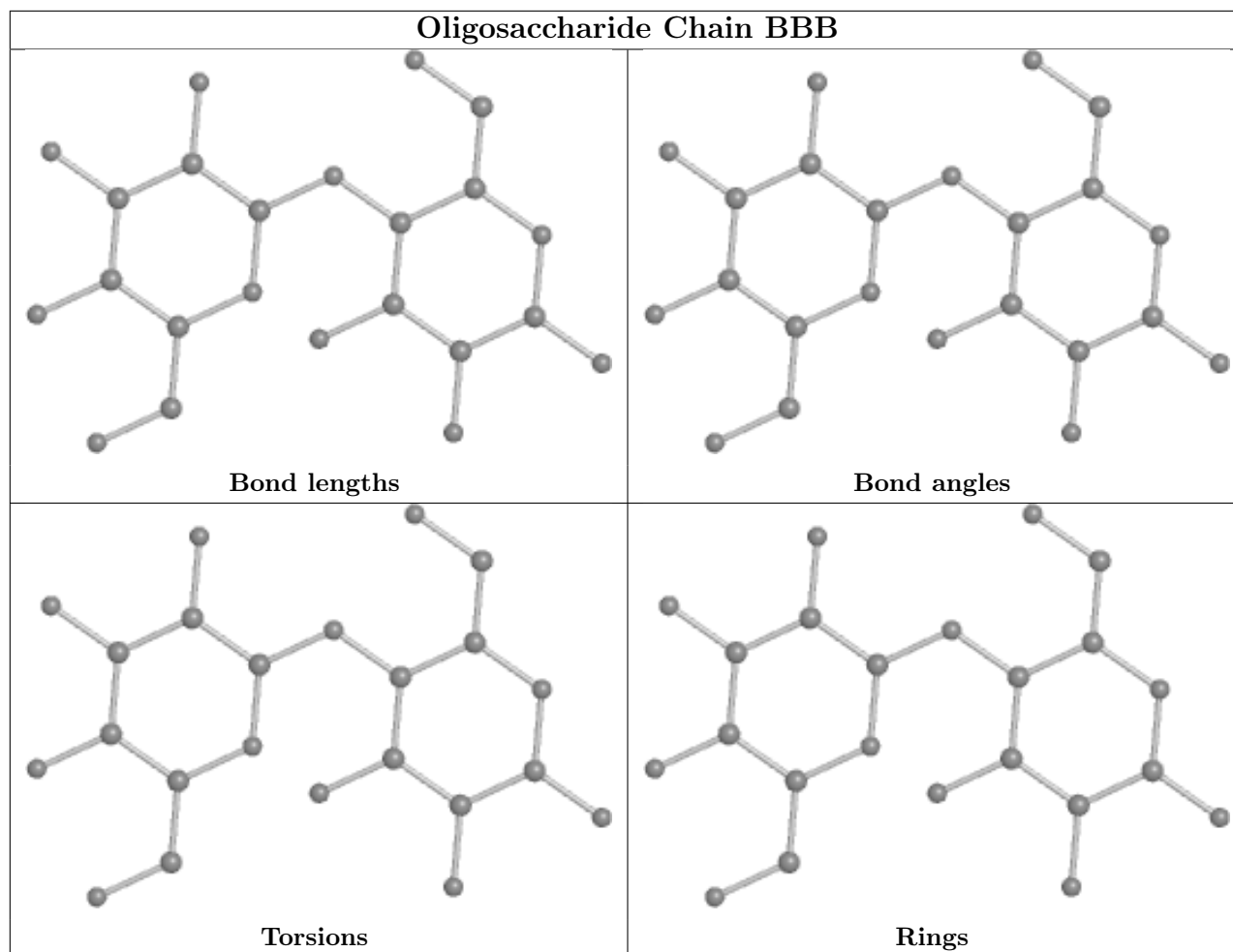
All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	BBB	1	BGC	O5-C5-C6-O6
2	BBB	1	BGC	C4-C5-C6-O6
2	BBB	2	BGC	C4-C5-C6-O6
2	BBB	2	BGC	O5-C5-C6-O6

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](#)

Of 8 ligands modelled in this entry, 3 are monoatomic - leaving 5 for Mogul analysis.

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	308	-	3,3,3	0.23	0	2,2,2	0.62	0
4	PEG	AAA	304	-	6,6,6	0.20	0	5,5,5	0.23	0
5	EDO	AAA	307	-	3,3,3	0.12	0	2,2,2	0.38	0
4	PEG	AAA	305	-	6,6,6	0.18	0	5,5,5	0.15	0
5	EDO	AAA	306	-	3,3,3	0.07	0	2,2,2	0.09	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	308	-	-	1/1/1/1	-
4	PEG	AAA	304	-	-	2/4/4/4	-
5	EDO	AAA	307	-	-	1/1/1/1	-
4	PEG	AAA	305	-	-	2/4/4/4	-
5	EDO	AAA	306	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	AAA	304	PEG	O2-C3-C4-O4
4	AAA	305	PEG	O2-C3-C4-O4
5	AAA	308	EDO	O1-C1-C2-O2
5	AAA	306	EDO	O1-C1-C2-O2
5	AAA	307	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

## 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	195/215 (90%)	-0.13	6 (3%) 49 58	32, 47, 82, 173	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	211	VAL	5.0
1	AAA	210	VAL	4.9
1	AAA	212	PRO	3.3
1	AAA	209	ALA	2.8
1	AAA	213	ILE	2.7

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

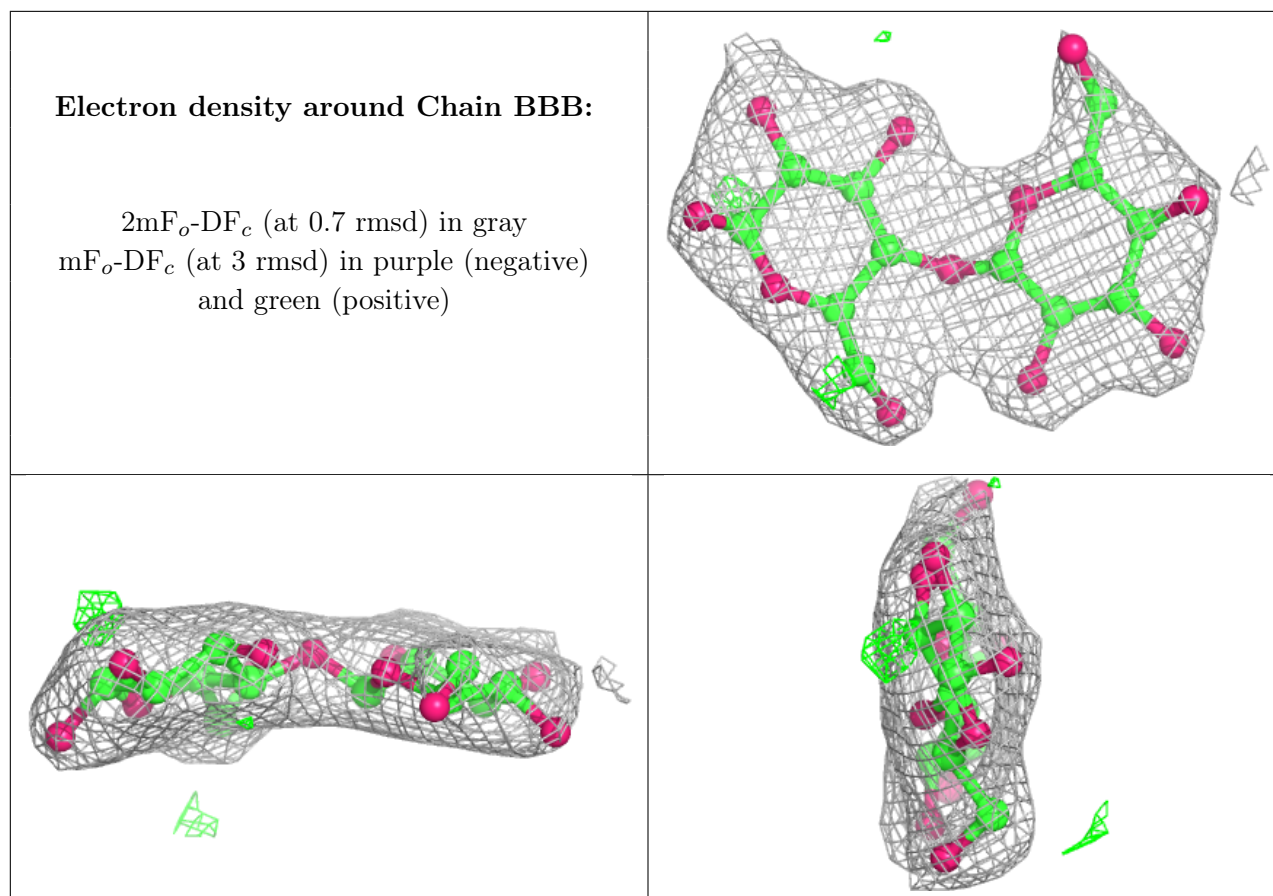
There are no non-standard protein/DNA/RNA residues in this entry.

### 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	BGC	BBB	2	11/12	0.92	0.12	66,74,84,89	0
2	BGC	BBB	1	12/12	0.96	0.09	49,59,67,72	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



## 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(Å <sup>2</sup> )	Q<0.9
5	EDO	AAA	308	4/4	0.74	0.24	77,79,86,90	0
4	PEG	AAA	305	7/7	0.75	0.22	80,84,90,95	0
5	EDO	AAA	306	4/4	0.82	0.41	73,75,77,78	0
5	EDO	AAA	307	4/4	0.86	0.13	85,86,87,93	0
3	CA	AAA	303	1/1	0.87	0.18	88,88,88,88	0
4	PEG	AAA	304	7/7	0.90	0.12	73,81,91,93	0
3	CA	AAA	302	1/1	0.96	0.03	49,49,49,49	1
3	CA	AAA	301	1/1	0.99	0.12	33,33,33,33	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.