WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 10, 2023 – 02:58 AM EDT

PDB ID	:	7MKS
Title	:	Crystal structure of the GH12 domain from Acidothermus cellulolyticus GuxA bound to cellobiose
Authors	:	Lunin, V.V.
Deposited on	:	2021-04-26
Resolution	:	1.85 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

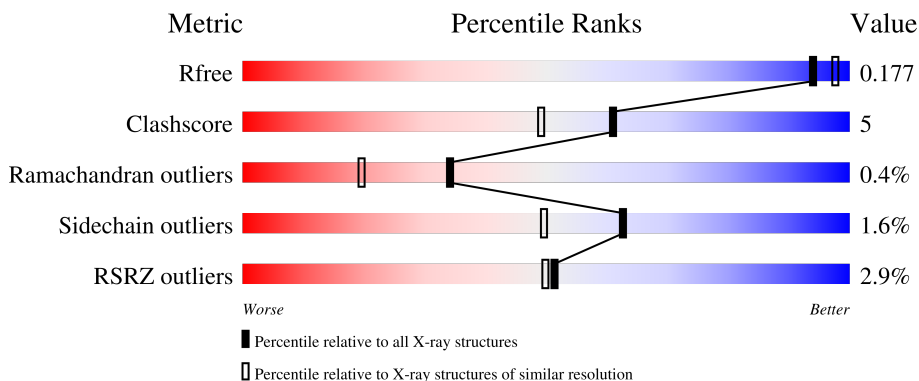|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance ⓘ

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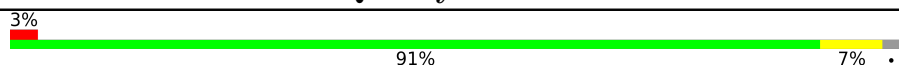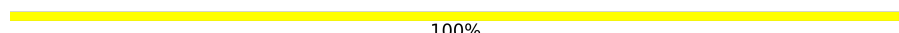*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.85 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
Rfree		0.177
Clashscore		5
Ramachandran outliers		0.4%
Sidechain outliers		1.6%
RSRZ outliers		2.9%

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))
R_{free}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	244	3% / 91% / 7% ·
2	A	2	100%

2   Entry composition ⓘ

There are 8 unique types of molecules in this entry. The entry contains 2271 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 Glycoside hydrolase, family 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	240	Total 1897	C 1172	N 326	O 390	S 9	0	18	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	237	ALA	-	expression tag	UNP A0LSH8
AAA	238	SER	-	expression tag	UNP A0LSH8
AAA	239	HIS	-	expression tag	UNP A0LSH8
AAA	240	HIS	-	expression tag	UNP A0LSH8
AAA	241	HIS	-	expression tag	UNP A0LSH8
AAA	242	HIS	-	expression tag	UNP A0LSH8
AAA	243	HIS	-	expression tag	UNP A0LSH8
AAA	244	HIS	-	expression tag	UNP A0LSH8

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

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

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

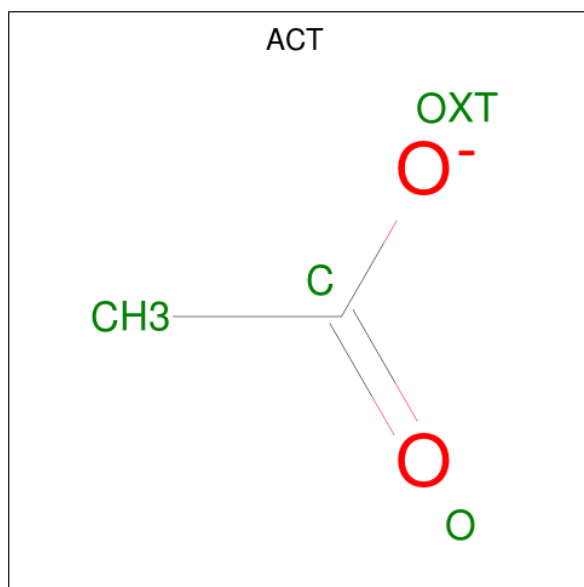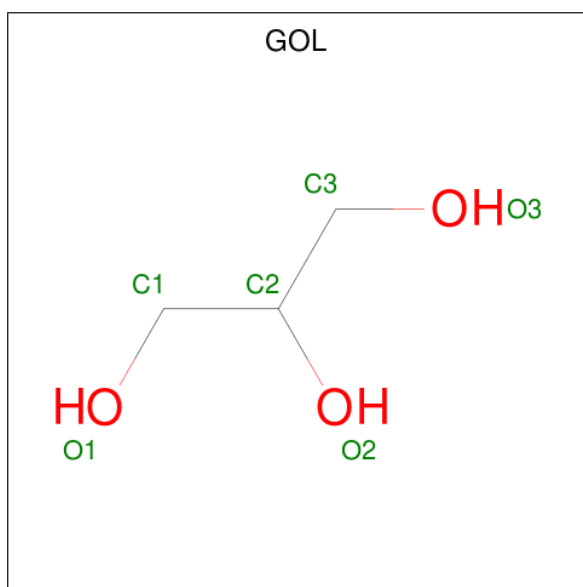| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	5	Total 5	Zn 5	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AAA	1	Total 1	Cl 1	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	AAA	2	Total 2	Na 2	0	0

- Molecule 6 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

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

- Molecule 7 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

C3

OH O3

C1   C2

HO

O1

OH

O2

| Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	AAA	1	Total 6	C 3	O 3	0	0

• Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	AAA	328	Total 329	O 329	0	1

W O R L D W I D E

PDB

PROTEIN DATA BANK
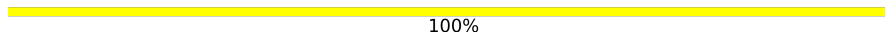
3 Residue-property plots ⓘ

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 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: Glycoside hydrolase, family 6

Chain AAA:
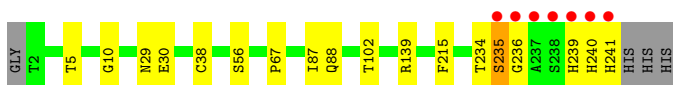
• Molecule 2: beta-D-glucopyranose-(1-4)-beta-D-glucopyranose

Chain A:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	66.23Å 66.23Å 127.18Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	34.10 – 1.85 34.09 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.6 (34.10-1.85) 99.6 (34.09-1.85)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.72 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.133 , 0.165 0.147 , 0.177	Depositor DCC
R_{free} test set	1419 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	12.4	Xtriage
Anisotropy	0.113	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 59.5	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.045 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2271	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.21% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: NA, ZN, CL, GOL, BGC, ACT

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	0.76	0/1943	0.88	1/2666 (0.0%)

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	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	139	ARG	NE-CZ-NH1	-6.32	117.14	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	234[A]	THR	Peptide

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	1897	0	1759	16	0
2	A	23	0	21	0	0
3	AAA	5	0	0	0	0
4	AAA	1	0	0	0	0
5	AAA	2	0	0	0	0
6	AAA	8	0	6	0	0
7	AAA	6	0	8	1	0
8	AAA	329	0	0	8	0
All	All	2271	0	1794	17	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 5.

The worst 5 of 17 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:235[B]:SER:N	8:AAA:404:HOH:O	1.89	0.91
1:AAA:87:ILE:HD12	1:AAA:235[B]:SER:HB3	1.69	0.72
1:AAA:241[A]:HIS:HD2	1:AAA:241[A]:HIS:O	1.74	0.71
1:AAA:88:GLN:OE1	8:AAA:405:HOH:O	2.11	0.68
1:AAA:102[A]:THR:HG22	8:AAA:561:HOH:O	1.95	0.65

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	255/244 (104%)	247 (97%)	7 (3%)	1 (0%)	34	19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	239	HIS

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	209/195 (107%)	204 (98%)	5 (2%)	49	33

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

Mol	Chain	Res	Type
1	AAA	56	SER
1	AAA	235[A]	SER
1	AAA	235[B]	SER
1	AAA	240[A]	HIS
1	AAA	240[B]	HIS

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	BGC	A	1	2	12,12,12	0.79	0	17,17,17	1.04	1 (5%)
2	BGC	A	2	2	11,11,12	1.39	2 (18%)	15,15,17	1.69	3 (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	BGC	A	1	2	-	0/2/22/22	0/1/1/1
2	BGC	A	2	2	-	0/2/19/22	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2	BGC	C2-C3	-2.44	1.48	1.52
2	A	2	BGC	O2-C2	-2.09	1.38	1.43

All (4) bond angle outliers are listed below:

| Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	2	BGC	C1-C2-C3	3.89	114.44	109.67
2	A	2	BGC	O5-C5-C6	3.22	112.25	107.20
2	A	2	BGC	O5-C1-C2	-2.26	107.28	110.77
2	A	1	BGC	O5-C1-C2	-2.02	106.69	110.28

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 A

Bond lengths

Bond angles

Torsions

Rings
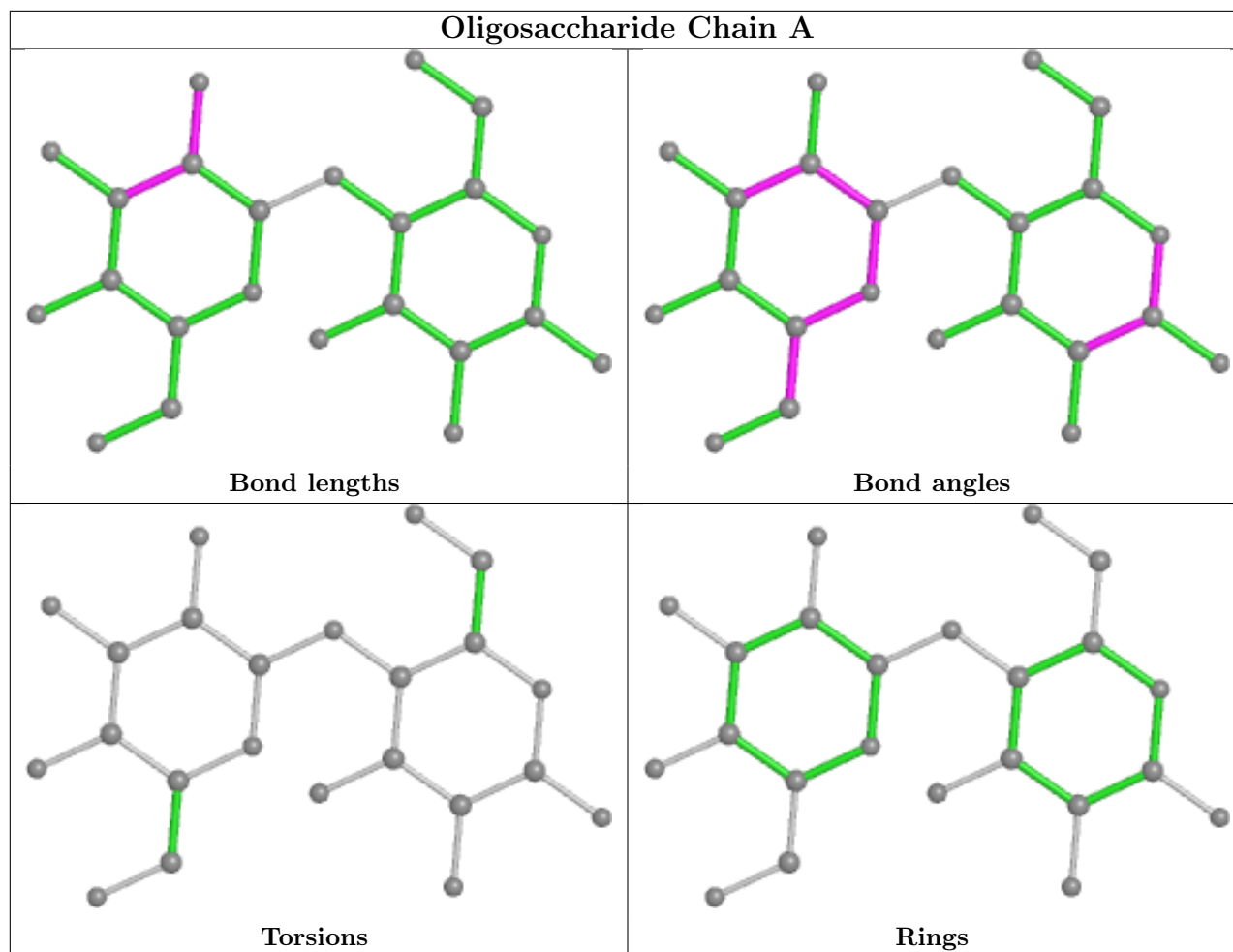
5.6 Ligand geometry ⓘ

Of 11 ligands modelled in this entry, 8 are monoatomic - leaving 3 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	$\#|Z| > 2$
6	ACT	AAA	309	5	3,3,3	0.77	0	3,3,3	0.85	0
7	GOL	AAA	311	-	5,5,5	0.19	0	5,5,5	0.51	0
6	ACT	AAA	310	3	3,3,3	1.19	0	3,3,3	0.75	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
7	GOL	AAA	311	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	AAA	311	GOL	1	0

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 ⓘ

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, 95th 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	240/244 (98%)	-0.56	7 (2%)	51	50	6, 11, 22, 62	0

The worst 5 of 7 RSRZ outliers are listed below:

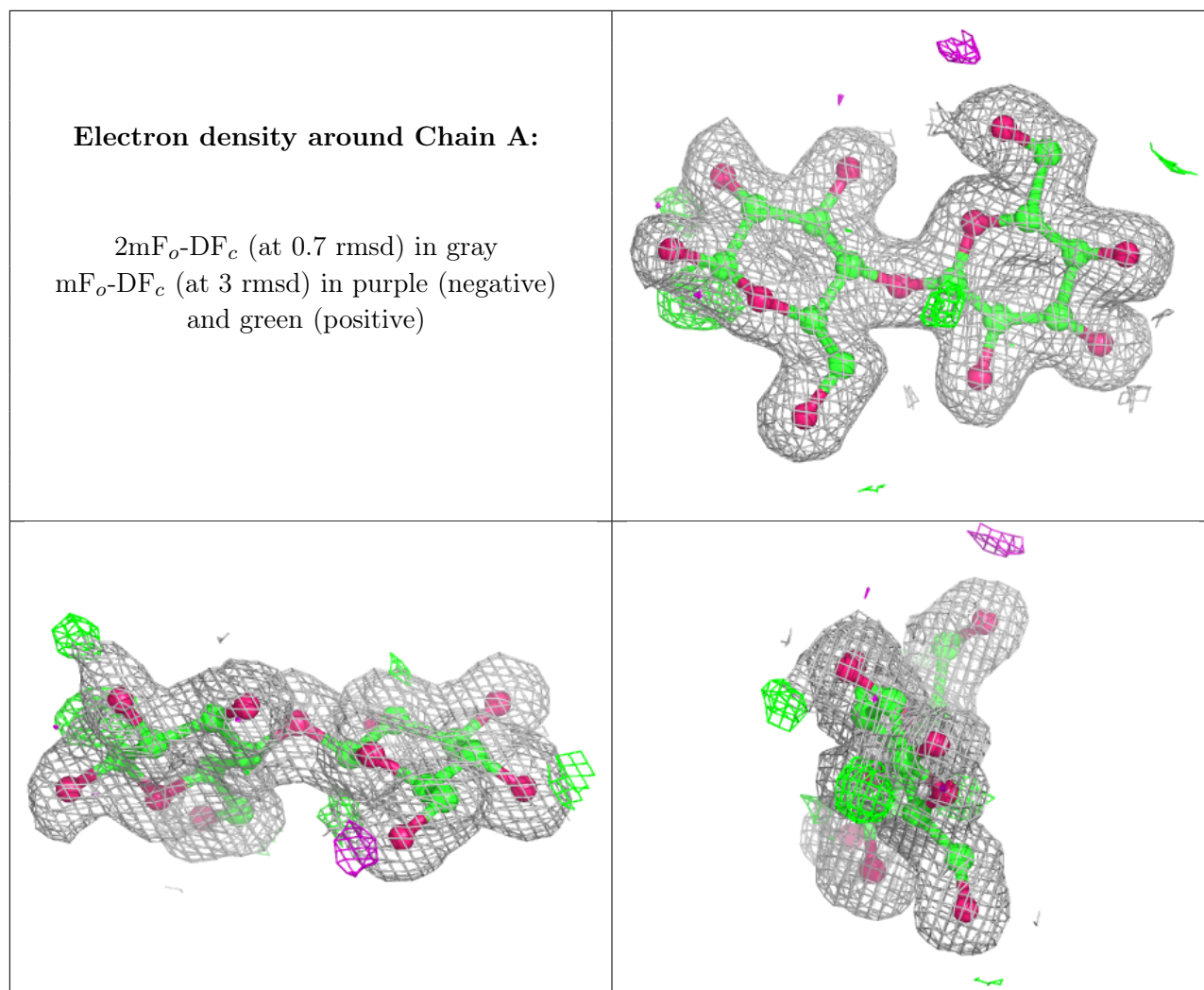| Mol	Chain	Res	Type	RSRZ
1	AAA	236[A]	GLY	5.4
1	AAA	240[A]	HIS	5.2
1	AAA	238	SER	3.6
1	AAA	235[A]	SER	3.3
1	AAA	239	HIS	2.5

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 ⓘ

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, 95th 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^2)	Q<0.9
2	BGC	A	1	12/12	0.96	0.07	9,12,19,23	0
2	BGC	A	2	11/12	0.98	0.08	9,9,10,10	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.

Electron density around Chain A:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

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, 95th 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^2)	Q<0.9
5	NA	AAA	308	1/1	0.94	0.08	25,25,25,25	0
3	ZN	AAA	304	1/1	0.95	0.14	48,48,48,48	1
7	GOL	AAA	311	6/6	0.95	0.18	18,20,21,22	6
6	ACT	AAA	310	4/4	0.96	0.15	19,24,26,30	0
6	ACT	AAA	309	4/4	0.96	0.12	24,32,36,40	0
5	NA	AAA	307	1/1	0.97	0.06	23,23,23,23	0
3	ZN	AAA	301	1/1	0.98	0.07	22,22,22,22	1
3	ZN	AAA	305	1/1	0.99	0.04	20,20,20,20	1
3	ZN	AAA	303	1/1	0.99	0.06	32,32,32,32	1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	CL	AAA	306	1/1	1.00	0.05	13,13,13,13	1
3	ZN	AAA	302	1/1	1.00	0.07	18,18,18,18	1

6.5 Other polymers ⓘ

There are no such residues in this entry.