



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 6, 2020 – 04:45 PM BST

PDB ID : 5II4  
Title : Crystal structure of red abalone VERL repeat 1 with linker at 2.0 Å resolution  
Authors : Sadat Al-Hosseini, H.; Raj, I.; Nishimura, K.; Jovine, L.  
Deposited on : 2016-03-01  
Resolution : 2.00 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 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 : 2.13.1  
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) : 2.13.1

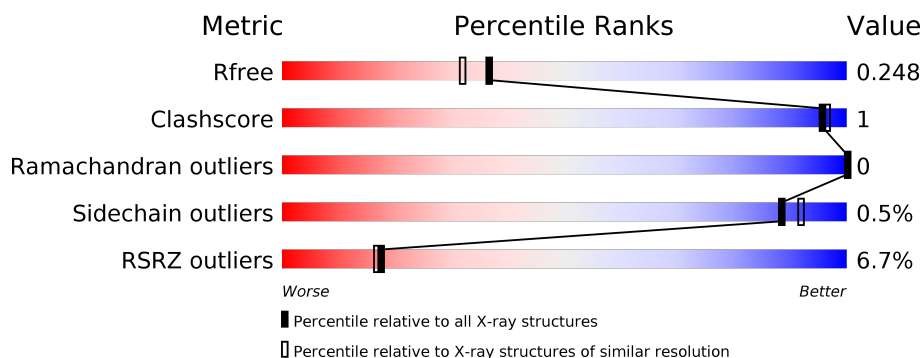
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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	A	517	<div> <div>6%</div> <div> <div></div> <div>90%</div> <div>•</div> <div>8%</div> </div> </div>
2	B	2	<div> <div>100%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7602 atoms, of which 3722 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 Maltose-binding periplasmic protein, Vitelline envelope sperm lysin receptor.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	477	Total	C	H	N	O	S	0	0	0
			7317	2369	3631	602	700	15			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3659	GLU	-	expression tag	UNP P0AEY0
A	3660	THR	-	expression tag	UNP P0AEY0
A	3661	GLY	-	expression tag	UNP P0AEY0
A	3662	HIS	-	expression tag	UNP P0AEY0
A	3663	HIS	-	expression tag	UNP P0AEY0
A	3664	HIS	-	expression tag	UNP P0AEY0
A	3665	HIS	-	expression tag	UNP P0AEY0
A	3666	HIS	-	expression tag	UNP P0AEY0
A	3667	HIS	-	expression tag	UNP P0AEY0
A	3669	THR	ILE	engineered mutation	UNP P0AEY0
A	3749	ALA	ASP	engineered mutation	UNP P0AEY0
A	3750	ALA	LYS	engineered mutation	UNP P0AEY0
A	3839	ALA	GLU	engineered mutation	UNP P0AEY0
A	3840	ALA	ASN	engineered mutation	UNP P0AEY0
A	3882	HIS	ALA	engineered mutation	UNP P0AEY0
A	3886	HIS	LYS	engineered mutation	UNP P0AEY0
A	3906	ALA	LYS	engineered mutation	UNP P0AEY0
A	3979	VAL	ALA	engineered mutation	UNP P0AEY0
A	3984	VAL	ILE	engineered mutation	UNP P0AEY0
A	4026	ALA	GLU	engineered mutation	UNP P0AEY0
A	4029	ALA	LYS	engineered mutation	UNP P0AEY0
A	4030	ALA	ASP	engineered mutation	UNP P0AEY0
A	4034	ASN	ARG	engineered mutation	UNP P0AEY0
A	4035	ALA	-	linker	UNP P0AEY0
A	4036	ALA	-	linker	UNP P0AEY0
A	4037	ALA	-	linker	UNP P0AEY0

*Continued on next page...*

Continued from previous page...

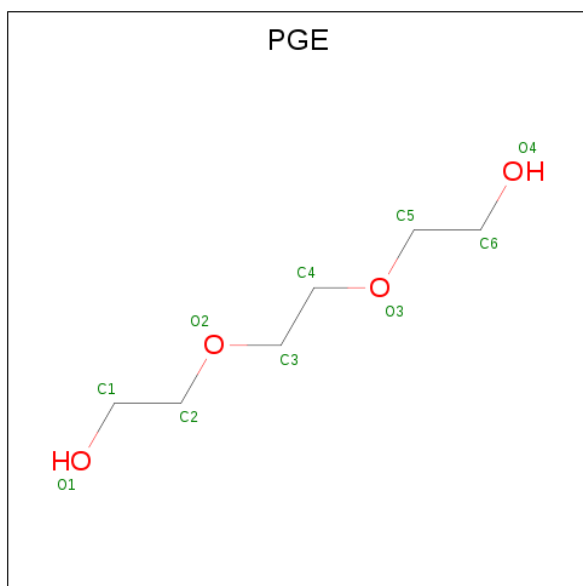
Chain	Residue	Modelled	Actual	Comment	Reference
A	4115	GLN	ASN	engineered mutation	UNP Q8WR62
A	4122	THR	ASN	engineered mutation	UNP Q8WR62
A	4142	TYR	ASN	engineered mutation	UNP Q8WR62
A	4171	GLN	ASN	engineered mutation	UNP Q8WR62

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



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	2	Total	C	H	O	0	0	0
			44	12	21	11			

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			24	6	14	4		
3	A	1	Total	C	H	O	0	0
			24	6	14	4		
3	A	1	Total	C	H	O	0	0
			24	6	14	4		
3	A	1	Total	C	H	O	0	0
			24	6	14	4		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			24	6	14	4		

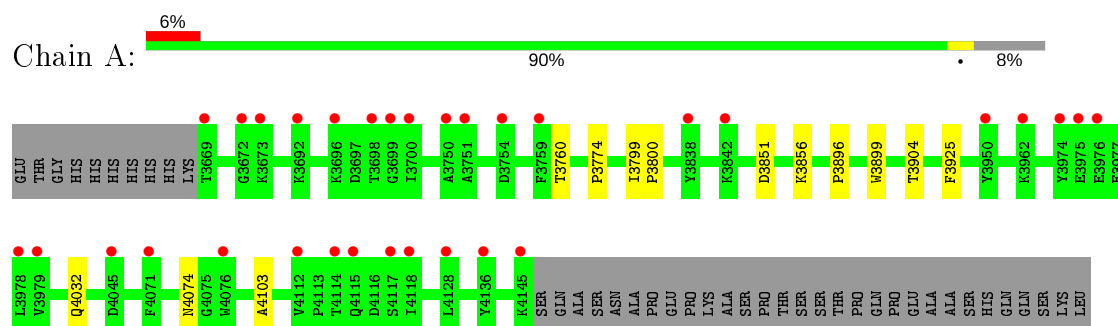
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	121	Total	O	0	0
			121	121		

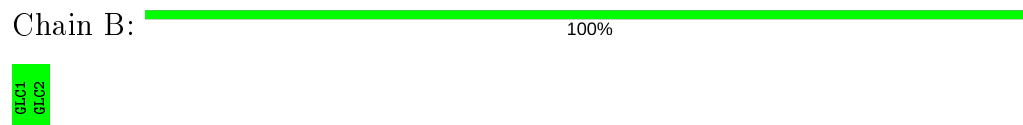
### 3 Residue-property plots [i](#)

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: Maltose-binding periplasmic protein,Vitelline envelope sperm lysin receptor



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.12Å 107.12Å 195.85Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.99 – 2.00 46.99 – 2.00	Depositor EDS
% Data completeness (in resolution range)	97.8 (46.99-2.00) 92.3 (46.99-2.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.82 (at 2.00Å)	Xtriage
Refinement program	PHENIX dev-1894_1692	Depositor
R, $R_{free}$	0.218 , 0.247 0.221 , 0.248	Depositor DCC
$R_{free}$ test set	2291 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	32.0	Xtriage
Anisotropy	0.507	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 53.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7602	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.80% 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: GLC, PGE

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	0.32	0/3779	0.45	0/5143

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	A	3686	3631	3631	8	0
2	B	23	21	21	0	0
3	A	50	70	70	0	0
4	A	121	0	0	0	0
All	All	3880	3722	3722	8	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 1.

The worst 5 of 8 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:3851:ASP:O	1:A:3856:LYS:NZ	2.31	0.62
1:A:3851:ASP:OD1	1:A:3856:LYS:NZ	2.34	0.60
1:A:4074:ASN:OD1	1:A:4074:ASN:N	2.39	0.55
1:A:4074:ASN:ND2	1:A:4103:ALA:O	2.36	0.53
1:A:3896:PRO:HA	1:A:3899:TRP:CE2	2.45	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	A	475/517 (92%)	471 (99%)	4 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	A	385/419 (92%)	383 (100%)	2 (0%)	88	92

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3904	THR
1	A	3925	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	3882	HIS
1	A	3886	HIS

### 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	GLC	B	1	2	12,12,12	0.54	0	17,17,17	0.67	0
2	GLC	B	2	2	11,11,12	0.67	0	15,15,17	0.62	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	GLC	B	1	2	-	0/2/22/22	0/1/1/1
2	GLC	B	2	2	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

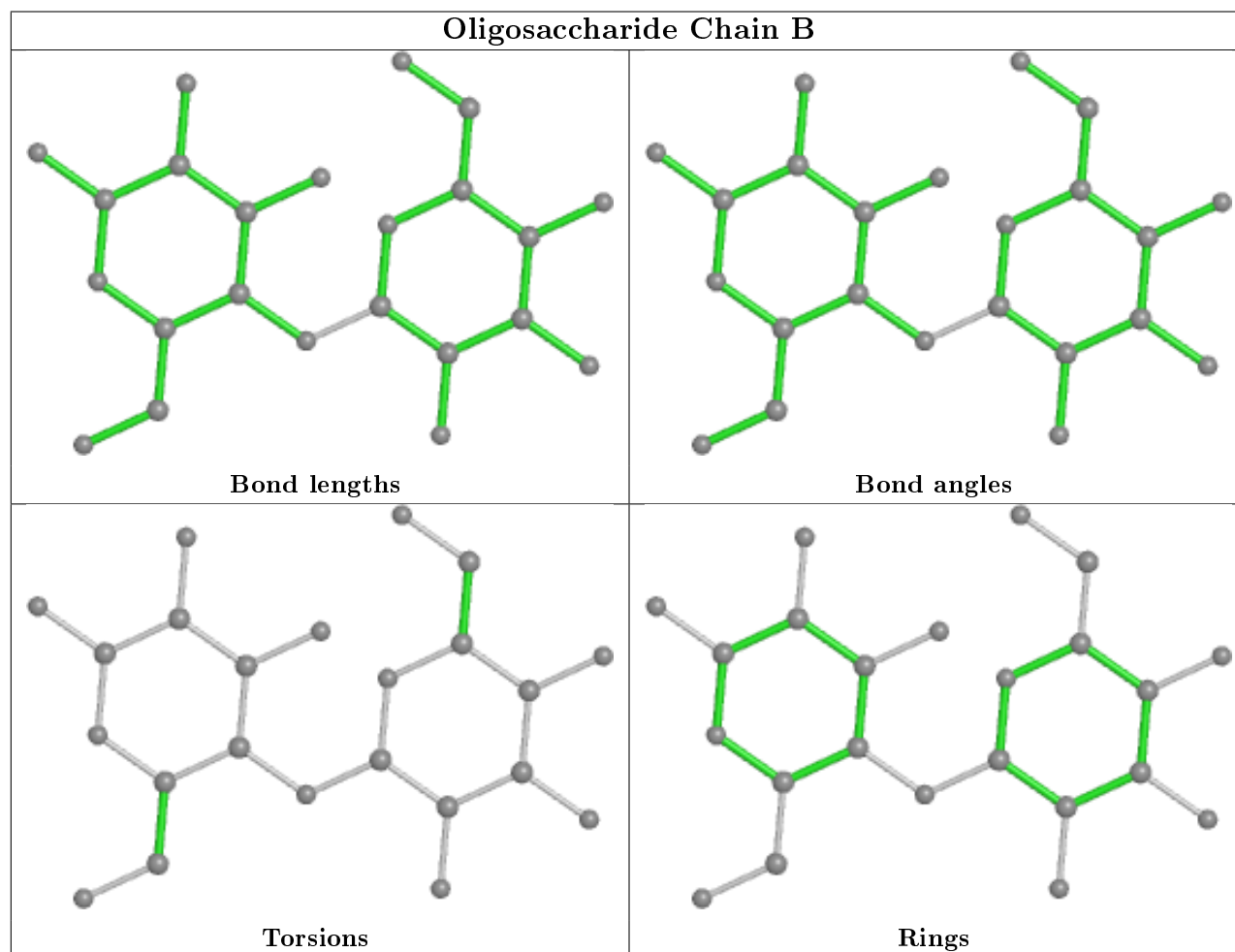
There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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
3	PGE	A	4206	-	9,9,9	0.29	0	8,8,8	0.49	0
3	PGE	A	4203	-	9,9,9	0.30	0	8,8,8	0.34	0
3	PGE	A	4202	-	9,9,9	0.32	0	8,8,8	0.34	0
3	PGE	A	4205	-	9,9,9	0.28	0	8,8,8	0.33	0
3	PGE	A	4204	-	9,9,9	0.31	0	8,8,8	0.31	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
3	PGE	A	4206	-	-	6/7/7/7	-
3	PGE	A	4203	-	-	1/7/7/7	-
3	PGE	A	4202	-	-	0/7/7/7	-
3	PGE	A	4205	-	-	3/7/7/7	-
3	PGE	A	4204	-	-	4/7/7/7	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	4206	PGE	O1-C1-C2-O2
3	A	4204	PGE	O1-C1-C2-O2
3	A	4204	PGE	C1-C2-O2-C3
3	A	4205	PGE	O3-C5-C6-O4
3	A	4204	PGE	O2-C3-C4-O3

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [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	477/517 (92%)	0.38	32 (6%) 17 17	28, 46, 89, 119	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	4114	THR	4.3
1	A	4045	ASP	3.6
1	A	3673	LYS	3.5
1	A	4118	ILE	3.5
1	A	3979	VAL	3.3

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

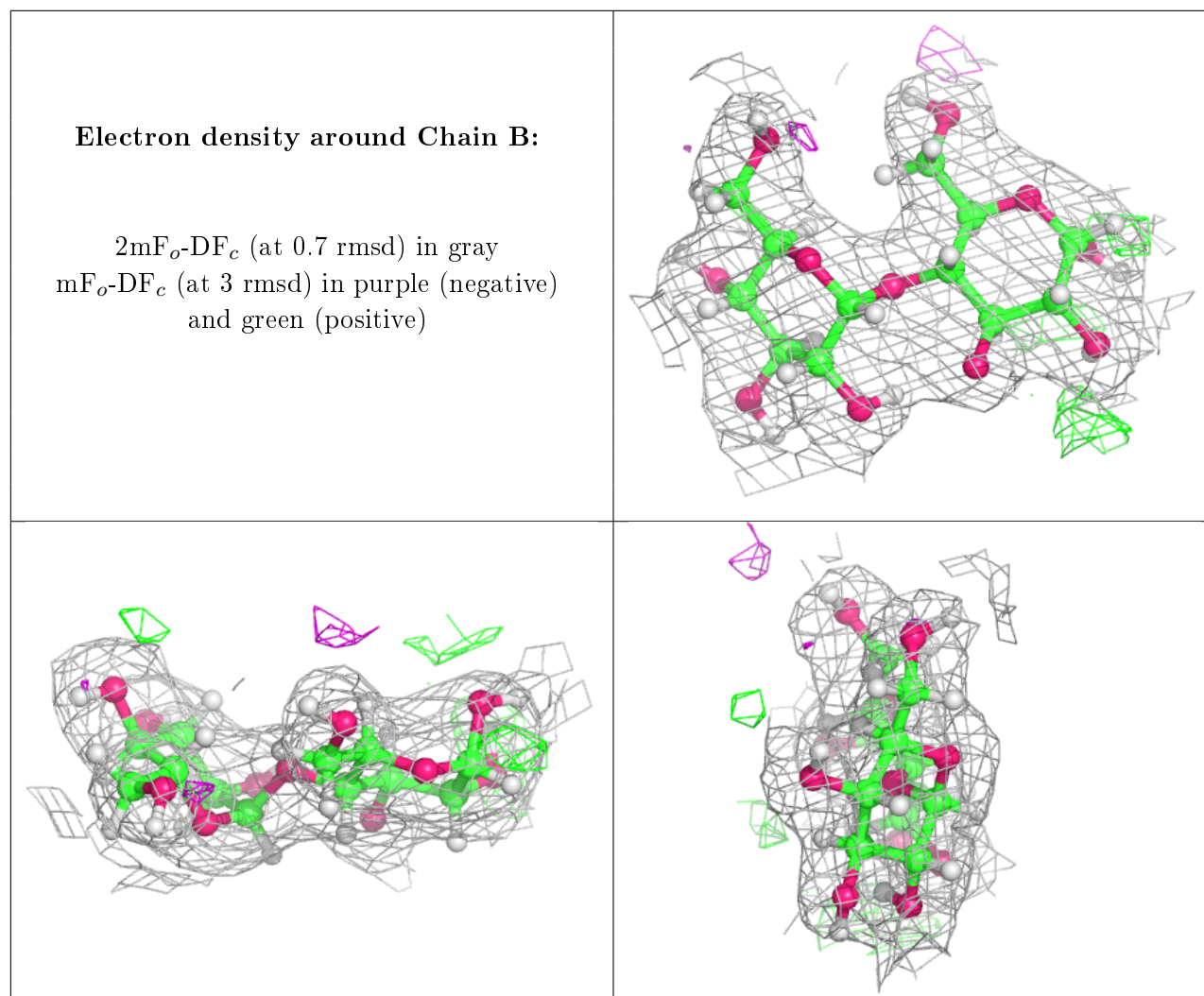
There are no non-standard protein/DNA/RNA residues in this entry.

### 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	GLC	B	1	12/12	0.95	0.16	31,37,50,58	0
2	GLC	B	2	11/12	0.97	0.13	30,34,38,41	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 ⓘ

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
3	PGE	A	4202	10/10	0.75	0.29	58,71,85,87	0
3	PGE	A	4203	10/10	0.77	0.26	59,71,88,90	0
3	PGE	A	4206	10/10	0.85	0.15	41,61,74,78	0
3	PGE	A	4204	10/10	0.85	0.16	58,70,78,78	0
3	PGE	A	4205	10/10	0.91	0.20	52,63,76,78	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.