Full wwPDB X-ray Structure Validation Report (i)

Feb 18, 2024 – 08:36 AM EST

PDB ID	:	4EGC
Title	:	Crystal Structure of MBP-fused Human Six1 Bound to Human Eya2 Eya Domain
Authors	:	Zhao, R.; Patrick, A.N.
Deposited on	:	2012-03-30
Resolution	:	1.99 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*
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:

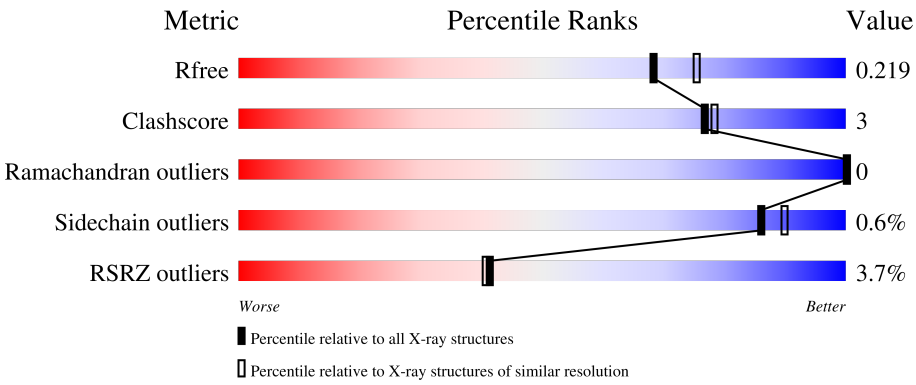
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.99 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 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	A	559	4% 90% 7% .
2	B	294	3% 79% 8% 13%
3	C	2	100%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7131 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 Maltose-binding periplasmic protein, Homeobox protein SIX1 chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	539	4242	2723	732	777	10	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	172	ALA	GLU	engineered mutation	UNP P0AEX9
A	173	ALA	ASN	engineered mutation	UNP P0AEX9
A	359	ALA	GLU	engineered mutation	UNP P0AEX9
A	362	ALA	LYS	engineered mutation	UNP P0AEX9
A	363	ALA	ASP	engineered mutation	UNP P0AEX9
A	367	ASN	-	expression tag	UNP P0AEX9
A	368	ALA	-	expression tag	UNP P0AEX9
A	369	ALA	-	expression tag	UNP P0AEX9
A	370	ALA	-	expression tag	UNP P0AEX9

- Molecule 2 is a protein called Eyes absent homolog 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	255	2053	1308	350	384	11	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	245	GLY	-	expression tag	UNP O00167
B	246	PRO	-	expression tag	UNP O00167
B	247	LEU	-	expression tag	UNP O00167
B	248	GLY	-	expression tag	UNP O00167
B	249	SER	-	expression tag	UNP O00167
B	250	PRO	-	expression tag	UNP O00167

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	251	GLU	-	expression tag	UNP O00167
B	252	PHE	-	expression tag	UNP O00167

- Molecule 3 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.

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

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
4	B	1	1	1	0	0

- Molecule 5 is water.

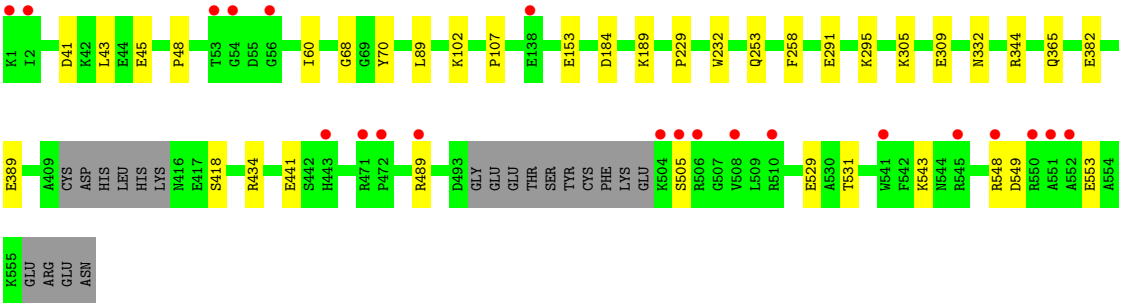
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	550	550	550	0	0
			Total	O		
5	B	262	262	262	0	0

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.

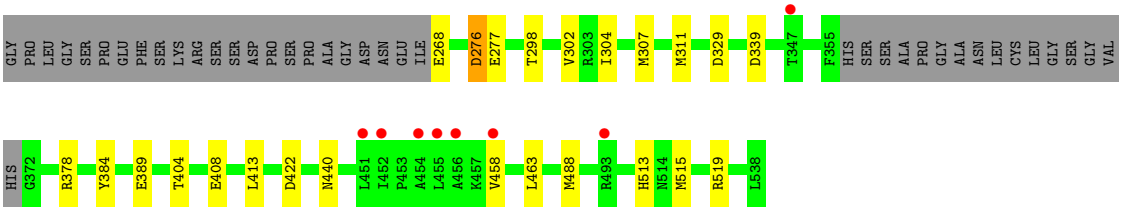
• Molecule 1: Maltose-binding periplasmic protein, Homeobox protein SIX1 chimera

Chain A:



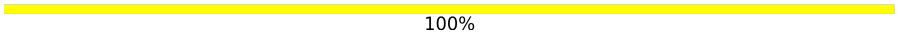
• Molecule 2: Eyes absent homolog 2

Chain B:



• Molecule 3: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	123.21Å 150.18Å 53.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.03 – 1.99 36.03 – 1.99	Depositor EDS
% Data completeness (in resolution range)	99.5 (36.03-1.99) 99.5 (36.03-1.99)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.51 (at 2.00Å)	Xtriage
Refinement program	PHENIX 1.7.3_928	Depositor
R, R_{free}	0.184 , 0.224 0.180 , 0.219	Depositor DCC
R_{free} test set	3488 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	26.3	Xtriage
Anisotropy	0.335	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 45.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7131	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% 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: MG, GLC

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	0.36	0/4348	0.50	0/5892
2	B	0.38	0/2095	0.54	0/2834
All	All	0.37	0/6443	0.51	0/8726

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	4242	0	4227	24	0
2	B	2053	0	2020	15	0
3	C	23	0	21	0	0
4	B	1	0	0	0	0
5	A	550	0	0	6	0
5	B	262	0	0	5	0
All	All	7131	0	6268	38	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 3.

All (38) 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:184:ASP:HB2	1:A:365:GLN:HB2	1.72	0.70
1:A:43:LEU:HD21	1:A:60:ILE:HD11	1.74	0.70
1:A:543:LYS:NZ	5:A:1204:HOH:O	2.27	0.67
1:A:382:GLU:OE1	2:B:519:ARG:NH1	2.30	0.64
2:B:277:GLU:HG2	5:B:705:HOH:O	2.00	0.61
1:A:489:ARG:HH22	1:A:548:ARG:HG2	1.67	0.58
1:A:549:ASP:O	1:A:553:GLU:HG2	2.05	0.57
1:A:489:ARG:NH2	1:A:548:ARG:HG2	2.21	0.56
1:A:189:LYS:NZ	5:A:1193:HOH:O	2.38	0.55
2:B:268:GLU:HB2	2:B:440:ASN:HB3	1.88	0.55
1:A:89:LEU:HD23	1:A:107:PRO:HG2	1.89	0.54
1:A:305:LYS:O	1:A:309:GLU:HG2	2.07	0.53
1:A:505:SER:HB2	1:A:531:THR:HB	1.92	0.52
1:A:41:ASP:OD2	1:A:434:ARG:NH2	2.44	0.50
2:B:339:ASP:OD2	2:B:378:ARG:HD3	2.12	0.49
2:B:422:ASP:OD2	5:B:867:HOH:O	2.20	0.49
1:A:45:GLU:O	1:A:48:PRO:HD2	2.13	0.48
1:A:389:GLU:OE1	5:A:931:HOH:O	2.20	0.47
1:A:229:PRO:HA	1:A:232:TRP:CE2	2.50	0.47
1:A:253:GLN:OE1	5:A:1195:HOH:O	2.20	0.47
1:A:102:LYS:NZ	5:A:1089:HOH:O	2.49	0.46
2:B:389:GLU:OE2	5:B:802:HOH:O	2.21	0.45
1:A:441:GLU:OE1	5:A:1146:HOH:O	2.21	0.45
2:B:513:HIS:HB2	2:B:515:MET:HE2	1.99	0.45
1:A:291:GLU:OE2	1:A:295:LYS:HE3	2.17	0.45
1:A:184:ASP:O	1:A:189:LYS:HE2	2.18	0.44
2:B:298:THR:O	2:B:302:VAL:HG23	2.17	0.44
2:B:458:VAL:HG13	2:B:463:LEU:HB2	1.99	0.44
2:B:307:MET:O	2:B:311:MET:HG3	2.18	0.43
2:B:404:THR:HG23	2:B:408:GLU:OE1	2.19	0.43
2:B:488:MET:HB2	2:B:488:MET:HE3	1.87	0.42
2:B:276:ASP:HB3	5:B:705:HOH:O	2.18	0.42
1:A:68:GLY:HA3	1:A:332:ASN:O	2.20	0.42
1:A:389:GLU:HG3	5:B:821:HOH:O	2.18	0.42
2:B:304:ILE:HG23	2:B:413:LEU:HD11	2.03	0.41
1:A:48:PRO:HG3	1:A:70:TYR:CE1	2.57	0.40
2:B:329:ASP:HA	2:B:384:TYR:OH	2.21	0.40
1:A:153:GLU:CD	1:A:344:ARG:HH12	2.25	0.40

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	533/559 (95%)	527 (99%)	6 (1%)	0	100	100
2	B	251/294 (85%)	242 (96%)	9 (4%)	0	100	100
All	All	784/853 (92%)	769 (98%)	15 (2%)	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	434/453 (96%)	431 (99%)	3 (1%)	84	88
2	B	223/253 (88%)	222 (100%)	1 (0%)	91	93
All	All	657/706 (93%)	653 (99%)	4 (1%)	86	90

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	258	PHE
1	A	418	SER
1	A	529	GLU
2	B	276	ASP

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	$\#\|Z\| > 2$
3	GLC	C	1	3	12,12,12	0.45	0	17,17,17	1.22	2 (11%)
3	GLC	C	2	3	11,11,12	0.75	0	15,15,17	0.91	1 (6%)

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	GLC	C	1	3	-	0/2/22/22	0/1/1/1
3	GLC	C	2	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	1	GLC	O5-C1-C2	3.08	115.78	110.28
3	C	1	GLC	C1-O5-C5	2.60	118.58	113.66
3	C	2	GLC	C1-O5-C5	2.24	115.22	112.19

There are no chirality outliers.

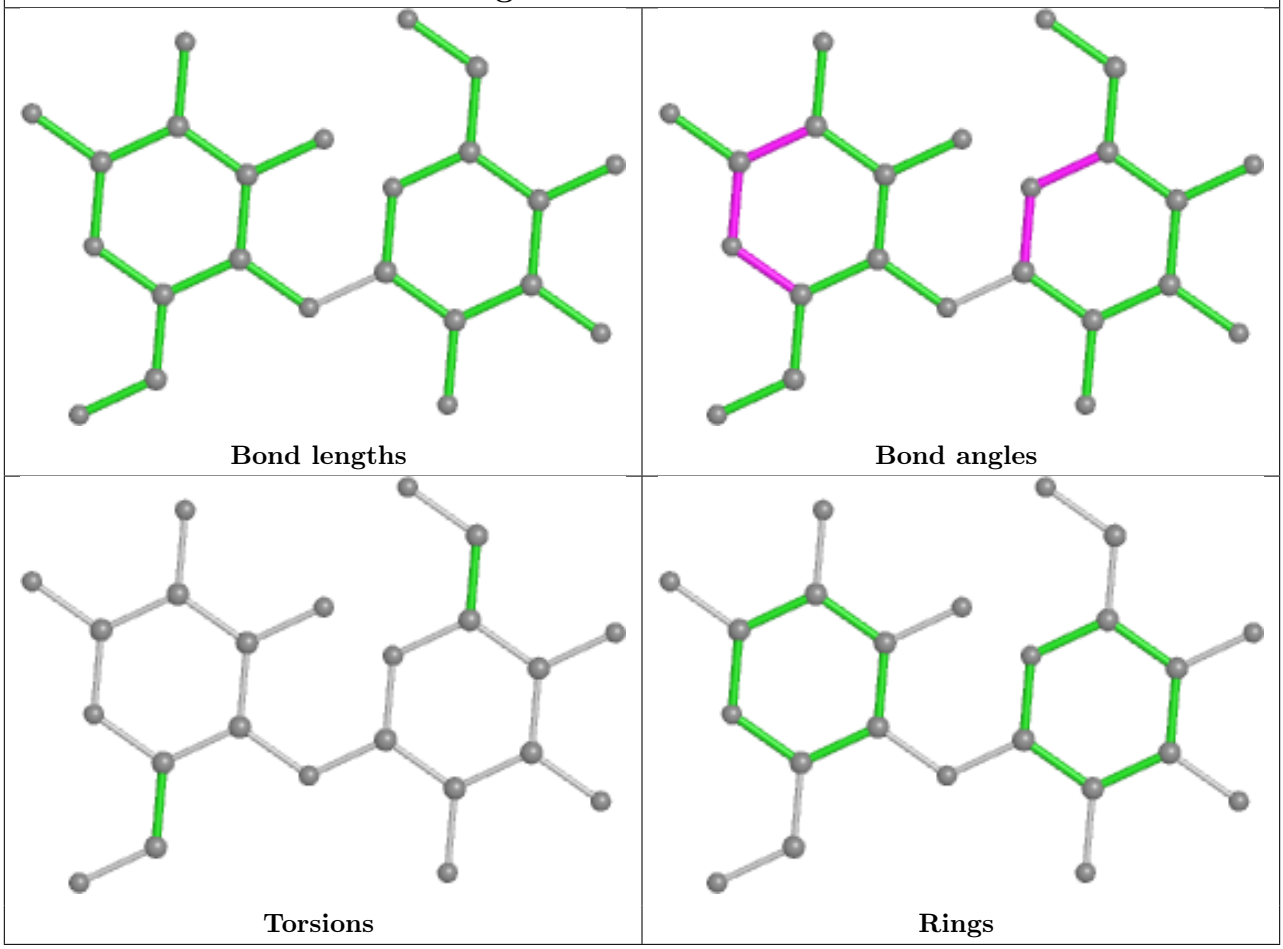
There are no torsion outliers.

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.

Oligosaccharide Chain C



5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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.

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, 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	A	539/559 (96%)	0.09	21 (3%)	39	38	16, 28, 50, 83	0
2	B	255/294 (86%)	0.10	8 (3%)	49	48	17, 26, 46, 64	0
All	All	794/853 (93%)	0.10	29 (3%)	41	41	16, 28, 49, 83	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2	ILE	4.6
1	A	504	LYS	4.3
1	A	1	LYS	4.2
1	A	54	GLY	4.2
1	A	548	ARG	4.1
1	A	508	VAL	3.6
1	A	472	PRO	3.4
1	A	506	ARG	2.9
2	B	451	LEU	2.9
2	B	493	ARG	2.8
2	B	347	THR	2.7
1	A	505	SER	2.6
1	A	489	ARG	2.6
1	A	541	TRP	2.6
1	A	550	ARG	2.5
1	A	471	ARG	2.5
2	B	455	LEU	2.5
1	A	138	GLU	2.5
1	A	56	GLY	2.4
2	B	456	ALA	2.4
1	A	510	ARG	2.3
2	B	458	VAL	2.3
1	A	53	THR	2.3
1	A	443	HIS	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	551	ALA	2.3
2	B	452	ILE	2.1
2	B	454	ALA	2.1
1	A	552	ALA	2.1
1	A	545	ARG	2.1

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)

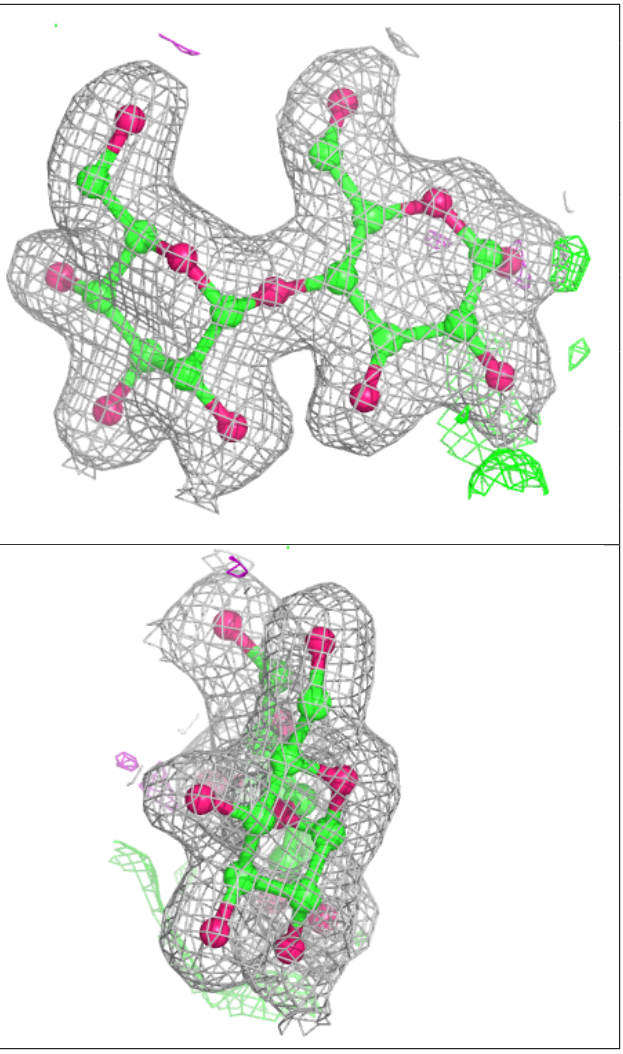
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	GLC	C	1	12/12	0.97	0.15	17,21,25,26	0
3	GLC	C	2	11/12	0.98	0.13	16,17,20,21	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 C:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-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
4	MG	B	601	1/1	0.97	0.21	31,31,31,31	0

6.5 Other polymers (i)

There are no such residues in this entry.